DOCUMENT RESUME

ED 450 255 CE 081 407

AUTHOR	Sticht, Thomas G.
TITLE	Testing and Accountability in Adult Literacy Education: Focus on Workplace Literacy Resources for Program Design, Assessment, Testing, & Evaluation.
INSTITUTION	Applied Behavioral and Cognitive Sciences, Inc., El Cajon, CA.
SPONS AGENCY	William and Flora Hewlett Foundation, Palo Alto, CA.; Work in America Inst., Scarsdale, NY.; Office of Vocational and Adult Education (ED), Washington, DC. Div. of Adult Education and Literacy.; National Inst. for Literacy, Washington, DC.
PUB DATE	1999-11-00
NOTE	89p.; Chapter 3 supported by THE CENTER/CCSD #54, Illinois.
AVAILABLE FROM	For full text: http://www.nald.ca/fulltext/sticht/testing/testing.pdf.
PUB TYPE	Opinion Papers (120) -- Reports - General (140)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	*Accountability; Adult Basic Education; Educational Finance; Educational Objectives; Evaluation Criteria; Federal Aid; Federal Government; Outcomes of Education; *Program Design; Program Effectiveness; *Program Evaluation; *Standardized Tests; Test Reviews; Test Use; *Workplace Literacy
IDENTIFIERS	Illinois

ABSTRACT

This report provides information on design and evaluation of workplace literacy programs (WLPs) to improve work force readiness and an overview of concepts about the nature, uses, and abuses of standardized tests in program evaluation and accountability. Chapter 1 takes a sociohistorical and sociopolitical perspective in discussing the knowledge and skills used by educational providers with differing philosophical views about what the goals of WLPs are and what kinds of WLPs should be developed to achieve them. Chapter 2 answers these four questions on WLP evaluation: By what criteria should a company judge its program's value? How should the government evaluate programs it funds? Are current government requirements for evaluation realistic and useful for companies receiving government funds? and On what basis should companies decide whether to fund WLPs? Chapter 3 illustrates how the Developing Organizational Effectiveness Through Employee Development approach to WLP evaluation outlined in Chapter 2 was used to evaluate national WLPs in the Chicago area. Chapter 4 addresses federal interests in standardized testing in adult education, the nature and uses of standardized tests, and special topics in their use. Chapter 5 illustrates the problem of determining how good is good enough in relation to workplace literacy in the context of the 1993 National Adult Literacy Survey. Reviews of eight tests used in adult basic education and English as a second language are appended. (YLB)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 450 255

Testing and Accountability in Adult Literacy Education

Focus on Workplace Literacy Resources for Program Design, Assessment, Testing, & Evaluation

Thomas G. Sticht

November 1999

Applied Behavioral & Cognitive Sciences, Inc.
2062 Valley View Blvd.
El Cajon, CA 92019-2059
(619) 444-9595

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

T. Sticht

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

CE081407

BEST COPY AVAILABLE

TABLE OF CONTENTS

Preface		**3**
Chapter 1	**Knowledge Resources for Designing and Delivering Workplace Literacy Programs**	**4**
Chapter 2	**Q & A on the Evaluation of Workplace Literacy Programs**	**17**
Chapter 3	**Case Study Using the "DO ED" Approach for Evaluating Workplace Literacy Programs**	**32**
Chapter 4	**Testing and Accountability in Adult Literacy Programs in the Workforce Education Act of 1998**	**51**
Chapter 5	**Determining How Many Adults Are Lacking in Workforce Literacy: The National and International Adult Literacy Surveys**	**69**
Appendix	**Reviews of Eight Tests Used in ABE & ESL**	**80**

Preface

Recent national and international surveys of adult literacy skills have raised questions about workforce readiness for international competitiveness. This report provides information on the design and evaluation of workplace literacy programs to improve workforce readiness, and an overview of concepts about the nature, uses and abuses of standardized tests in program evaluation and accountability. This is not a "how to do it" guidebook. Rather, it discusses concepts and issues and provides bibliographic resources for those readers who want to learn more about how to design, develop, and evaluate literacy programs in the workplace and other contexts.

Workplace literacy or basic skills programs are programs offered at a given workplace and generally are aimed at preparing employees for performing job-linked literacy and numeracy tasks, such as filling our requisition forms in a clerical position or preparing to learn statistical process control. However, much of the discussion is applicable to other types of programs for workforce education and lifelong learning, family literacy, academic literacy and other aspects of basic skills education (reading, writing, mathematics, English as a Second Language-ESL).

Materials in Chapters 1 and 2 were prepared with support from the Work in America Institute and the U. S. Department of Education, Division of Adult Education and Literacy. Chapter 3 was prepared under a contract with THE CENTER/CCSD #54, an adult education organization near Chicago, Illinois. The preparation of Chapter 4 was supported by a contract from the U. S. Department of Education, Division of Adult Education and Literacy, while Chapter 5 was prepared, in part, under a contract with the National Institute for Literacy (NIFL) to the San Diego Community College District.

Integration of these various papers into the present report was supported in part by a grant from the William and Flora Hewlett Foundation to the Applied Behavioral & Cognitive Sciences, Inc. The opinions, viewpoints, and positions stated in this report are those of the author and they do not necessarily reflect the policies or positions of any of the organizations named herein.

Chapter 1

Knowledge Resources for Designing and Delivering Workplace Literacy Programs

Designing and delivering workplace literacy programs are activities that take place within a system of values and beliefs about what the legitimate aims of such programs should be and how these aims can best be achieved. Many of these values and beliefs have been acquired by educators who have been involved in youth and adult education and job training programs for the last quarter of a century. During this time, many different providers have developed approaches to teaching basic skills to youth and adults for a variety of purposes, such as for completing the high school diploma, for obtaining job training and work, for reaching personal development goals (e.g., reading the Bible), and for social activism to enrich the lives of those living within a given community.

The knowledge gained through the historical experiences of those who have designed and delivered youth and adult literacy programs in the past, including workplace literacy programs, greatly influences how they go about the task of developing such programs today. It may be useful for those contemplating the introduction of workplace literacy programs to know about the beliefs and practices of various educational "providers" who may be engaged in the design and delivery of job-linked literacy (primarily the basic skills of reading, writing, and arithmetic) programs.

This chapter takes a sociohistorical and sociopolitical perspective in discussing the knowledge and skills that have been used by educational providers holding differing philosophical views about what the goals of such programs are (or should be) and what kinds of programs should be developed to achieve these goals.

The chapter first discusses the major national policy impetus of a quarter century ago for educational reform in the nation, the War on Poverty, and how that influenced the education and training of numerous educational providers regarding how adult literacy programs should be designed and delivered. The goal is to set the stage for understanding how the various approaches to job-linked literacy that exist today emerged from this historical background.

The chapter then discusses the shift in the national policy emphasis for educational reform that was announced in 1983 with the publication of the report on *A Nation at Risk.*. This shift has brought about the current interest in workplace literacy programs and, in general, a first-time emphasis upon the education and training of non- management, non-supervisory, "line" employees in America's workplaces. Interestingly, changes that have taken place in just the last three to five years, in various so-called "high performance" businesses and industries which emphasize the "empowerment" of line employees by having them participate in collaborative planning, decision making, and quality monitoring, seem to be influencing the processes for the design and delivery of workplace literacy programs.

The shift seems to be away from the "top-down" approach advocated by the U. S. Departments of Education and Labor in their report on *The Bottom Line*, in which a literacy task analysis or audit is performed and a curriculum is developed based on that task analysis. Under the "empowerment" philosophy, more and more businesses and educational providers are following an "interactive" approach in which educational providers encourage both management (top-down) and employees (bottom-up) to

participate interactively to determine what the workplace literacy program will look like and when and how it will be delivered.

From Poverty Warriors to World Competitors

In 1983 the report on *A Nation at Risk* asserted that America was losing its competitive edge in the world economic order. Because this report was paid for by the U.S. Department of Education, it is not surprising that this loss of our competitive edge was placed, to a very large extent, on the inadequacies of the U.S. education system in preparing our nation's workforce with the literacy, mathematics, science, and other, so-called "higher-order" skills needed to compete in the new world marketplace. This report was followed by a plea for reforms that would require greater funds for education.

The significance of the *Nation at Risk* report for workplace literacy is that it focussed on the role of education for making the nation's businesses and industries competitive in the "new world economic order." This was a change from the earlier major call for educational reform that focussed on problems of poverty. Indeed, the War on Poverty was a rallying cry of the major initiatives of the 1960's that led to the implementation of Head Start preschool programs, compensatory education in the public schools, the adult education act that institutionalized adult basic and secondary "remedial" or "second chance" education in the U.S. Department of Education, and extensive programs to get people off welfare, into jobs, and out of poverty (e.g., the Job Corps; the Manpower Development and Training Act [MDTA] and its successors (CETA-Comprehensive Employment and Training Act which became the JTPA-Job Training Partnership Act, and which is now incorporated into the Workforce Investment Act (WIA) of 1998).

In the War on Poverty, attention was focussed on the needs of individuals for education and training for work. This focus I call work*force* literacy because, for the most part, the aim was to upgrade the literacy and technical skills of those entering or already in the workforce who were not employed in a particular job field or workplace. The hope was that by providing individuals with education and training they would find jobs and work their way out of poverty.

Generally, programs for youth and adults (age 15 and above) were delivered in two main systems: the institutionalized, "second chance" system and the community-based organizations system.

The Institutionalized, "Second-Chance" System

This system was (and still is) comprised of two main subsystems: *The traditional adult school system* that has been around for decades in which adults attend classes in high schools in the evening or in community colleges at various times and in correctional institutions. In such programs the aim has generally been to help adults acquire the basic and secondary education needed to eventually obtain a high school diploma or equivalency certificate. *Government sponsored training programs* such as those found in the Job Corps, the welfare system (the early Work Incentive program; presently Job Opportunity and Basic Skills programs and Job Training Partnership Act programs) programs.

In these institutional second chance systems, literacy education typically focussed on improving individual's basic skills to the point that they could pass the General Education Development (GED) test battery or their general literacy skills were raised to a level (e.g., 8th grade level) that qualified them for job training and work.

In these programs, the aim was to render the least skilled members of the youth and adult workforce skilled enough to actually find, obtain, and retain work. But because the programs did not focus on a specific job, the literacy training tended to be "general." That is, it focussed on providing reading, writing, and arithmetic skills using the contents and methods of those related to progressing through the K-12 public school system. This is part of the reason many programs were referred to as "second chance" programs. Many youth and adults in the Job Corps and adult basic education programs had passed through the K-12 public system, or through 9-10 grades or so before dropping out, and had achieved only minimal levels of basic skills and had not obtained a high school diploma when they went through the K-12 system the first time. To a large extent, then, the youth and adult literacy programs of the War on Poverty were therefore considered as a second chance at learning basic skills for those who had failed to achieve well in the public school system the first time.

At times, literacy programs supplemented academic, general literacy education with "life skills," "functional literacy skills," or "real life skills." This typically involved teaching students to use basic skills in accomplishing tasks in the areas of consumer economics, occupational knowledge (e.g., how to read want ads), transportation (e.g., how to read a bus schedule); health (e.g., how to read medicine bottle labels) and so forth.

In some programs, such as the Job Corps, the replication of the K-12 concepts of content areas versus academic or basic skills development were maintained. In the Job Corps, youth attended academic (basic skills) education aimed eventually at obtaining the GED certificate for part of the day, and vocational training the other part of the day. This maintained the idea that basic skills or academic skills are something different from vocational skills. That reading, writing, and arithmetic are learned in one type of course (academic) and job skills in another (vocational education).

Many MDTA-sponsored programs followed a similar approach of providing remedial academic education part of the day, employability (how to make a resume, engage in a job interview, dress appropriately for work, etc.) training another part of the day, and job skills training at another time.

In these various "institutionalized" programs, the students generally had their skills assessed at entry into the program, and they were then placed in the curriculum. The latter typically followed an approach of beginning with the lowest level of skill in the area being taught, and then the materials in the curriculum progressively increased in difficulty as skill was developed.

For instance, for those students at the lowest levels of reading skill, basic decoding skills, including phonics and other "word attack" skills were taught, generally using some form of programmed instruction. Then numerous other skills were introduced sequentially to bring the student up to higher levels of reading comprehension. At around the 8th grade level of skill, students generally qualified for some form of job training, and instruction generally shifted to GED preparation for non-high school graduates.

Instructional Methods. While traditional primary and secondary school textbook and classroom techniques were followed by teachers in the federally and state funded adult basic education system (high school night schools for adults; community college based ABE), a general feature of many institutionalized programs, such as those in the Job Corps, many Work Incentive Program and other job-training settings, was that a pre-developed, fairly highly structured curriculum sequence, similar to programmed instruction, was developed and students were placed into it at their appropriate level of

skill. Students then performed numerous workbook activities or participated in classroom didactic activities and gradually progressed in their skills.

With the introduction of computers on a large scale, these educational activities and procedures were transferred onto computers. The PLATO system (today transmitted over NovaNet) offered a developmental sequence of tutorials and "skills and drills" in reading and math to adult learners in the early 1970s.The Job Corps incorporated computer managed instruction that kept track of the hundreds of proficiency checks that were used to determine if students had mastered one skill before proceeding to the next (the Job Corps approach has been converted for "civilian" use by U.S. Basics and has been marketed as the Comprehensive Competencies Program - CCP).

Community-Based Organizations

Supplementing the traditional institutional programs were numerous community-based programs. These were essentially of two main kinds, those initiated by the major voluntary literacy provider associations, Laubach Literacy and Literacy Volunteers of America, and numerous independent organizations that were started by concerned citizens.

The Laubach & LVA organizations were made-up of a national central office and numerous councils located throughout the nation. They focussed mainly on teaching the very basic decoding and word attack reading skills to youth and adults who were almost totally illiterate. Their major approach was to use thousands of volunteer tutors throughout the nation to accomplish what was captured in the slogan of Frank Laubach, "Each one teach one."

Each of the national volunteer organizations developed very structured teaching methods that could be followed by volunteer tutors anywhere in the nation. By following the teaching "script" the tutors could work in a one-on-one manner with a student and lead the student from illiteracy to literacy at about the fourth grade level.

Community-Based Organizations (CBOs) represent a wide diversity of literacy providers who have typically viewed literacy as a means to improving the social conditions of individuals and their communities. In the War on Poverty, the federal government's Community Action Program aimed at bringing about improvements in poor communities. As a part of these activities, community- based groups were formed that frequently incorporated literacy education as a part of their overall plan to bring about change in the community.

Religious charities, retired school teachers, political activists and just concerned citizens have formed literacy programs that operate out of storefronts, back rooms, out buildings, living rooms, and in some highly successful programs, out of specially designed facilities.

Some CBOs, such as the Center for Employment Training and the Urban League, have grown to become major institutions that offer literacy and job skills training to thousands of adults each year. They have learned to efficiently obtain the millions of federal and state dollars that are available for the education and training of disadvantaged youth and adults.

Early in the War on Poverty, CBOs were not included in the category of organizations that could receive federal and state funds for literacy programs. But over the years that has changed and many now participate almost as institutionalized members of the "second chance" system.

As mentioned, for a large number of the CBOs a major mission is to change the life circumstances of the poor. According to a report by the Association for Community Based Education (1984), CBO literacy programs seek "empowerment of the individual and development of the community" and they emphasize "learner-centered methodologies" and learner "participation, with teachers or tutors playing a facilitative rather than a didactic role."

The Poverty Warriors

The War on Poverty mobilized a broad segment of education "providers" in the provision of education to disadvantaged youth and adults. These educatonal "poverty warriors," though all aimed at improving the life circumstances of their students, differed remarkably in their philosophical beliefs about just what was being done, why, and how it should be done.

Working Within the System. The traditional "second chance" institutions and the major volunteer organizations (Laubach Literacy, LVA) focussed on teaching reading, writing, and mathematics to youth and adults who had failed in or been failed by the "first chance" educational system, and needed "remedial education" to acquire literacy and numeracy as instrumental skills for individual personal and economic growth.

In these programs, the teachers or tutors focussed on making the *individual learner* more broadly competent in the traditional academic skills of reading, writing, and arithmetic, with some additional attention to "life skills" at times. The general vision was to improve the individual's ability to work better within the present sociopolitical system, to cope with the exigencies of life, to help the person take charge of and improve the social circumstances of her or his personal life, the lives of the person's family members, and the overall social circumstances of the community in which students lived.

This vision included the important goal of improving the person's personal competitiveness in the workforce by providing education to improve skills and, importantly, to obtain the valued high school diploma equivalency certificate which was the "ticket" to employment in many businesses. In short, in the "second chance" system, literacy educators focussed on improving the *personal competitiveness of individuals* to help them raise themselves, their families, and their communities out of poverty.

Working to Change the System. In contrast to the large "second chance" system, a large number of CBOs were political activists and were stimulated by the writings of Paulo Freire in his *Pedagogy of the Oppressed* (1970). In the way of thinking of many of these activists, the problem of illiteracy resulted from the oppression of the poor by the management and governing classes. From this point of view, the sociopolitical system of the United States, with its industrial- and government-based "ruling" classes was viewed as oppressing the poor and restricting them from developing higher levels of literacy through lack of proper funding of education in poor neighborhoods, bias and discrimination in admitting the poor to job training and well paying jobs, and in general withholding of benefits in health, education, transportation, child care and so forth. The aim of this oppression was to provide a base of low paid, laborers upon whose backs large profits could be made by industrial giants and government benefactors.

For many educators in the CBOs, then, the goal of literacy education was not simply to prepare learners to take their obligatory positions in the existing social structure, in which the world of work was a major substructure, but, rather, to *empower* their learners to take action to change the existing power relationships and to bring about a more equitable

sharing of power among the poor and the better off, the governed and the government, the employer and the employee, and managers and workers.

The Shift from Personal Poverty to National Competitiveness as the Basis for Educational Reform and Workplace Literacy

The 1983 report on *A Nation at Risk* shifted the focus of concerns for educational reform from the personal competiveness and the plight of persons living in poverty, to the competitivenss of America's industries and businesses in the world marketplace. It argued that many of the problems faced by business and industry resulted from the low skills of many school graduates and the failure of business and industry to find sufficiently skilled workers.

Over the next few years a growing concern was expressed by various business and government leaders about the problems of America's international competitiveness, the poor performance of the public schools in educating students, and the need to upgrade the skills of the American workforce.

In 1987, the Hudson Institute released a landmark report sponsored by the U.S. Department of Labor that made the point that, even if school reform could be implemented right away, and made successful, this would not do much for the workforce of the year 2000 because two-thirds of that workforce is already on the job (Johnston & Packer, 1987, p. 75). The implication of this observation to the government was that more needed to be done to improve the skills of the present, employed workforce. This lead to the emergence of the *workplace literacy* programs of the U. S. Departments of Education and Labor.

Government funding of workplace literacy programs for the last several years has made the design and delivery of job-linked literacy programs a growth enterprise in the United States of America, as well as in some other industrialized nations (e.g., Canada; Britain) (Taylor, Lewe, & Draper, 1991).

The growth of funds for workplace literacy programs has produced a movement of traditional work*force* literacy or basic skills "providers" into the work*place* arena. This includes educational companies (e.g., Sylvan Learning Centers; Jostens Learning; IBM; Performance Plus; Computer Curriculum Corporation; U.S. Basics; etc.), community colleges, four year colleges and universities (e.g., Indiana University; City University of New York; Columbia University; University of California at Berkeley; etc.), secondary school districts with adult education departments (e.g., Los Angeles Unified School District; etc.), assorted vocational /technical, publicly funded and private schools, and numerous community-based organizations of both a local (e.g., Push Literacy Action Now [PLAN] in Washington, DC) and national scope (e.g., Laubach Literacy; Literacy Volunteers of America).

Added to this traditional array of adult literacy providers are a number of new providers entering from the field of organizational management and training consulting (e.g., professional associations such as the American Society for Training Development; American Banking Association; the National Alliance for Business; the Work in America Institute; etc.). Because federally funded, workplace literacy programs deal with current employees, workplace literacy efforts have been mounted by labor unions (e.g., United Auto Workers; AFL/CIO) as federally-funded additions to their traditional education efforts for union members.

Though the trends are not absolute, examination of a large number of workplace literacy programs suggests that many of those initiated by traditional "second chance" educators

and business/industry partnerships seem to emphasize the "top-down" approach in which management and educator teams determine the need for basic skills education, and design, develop and deliver the programs to employees.

On the other hand, "bottom up" approaches are more likely to be found in those programs that are initiated by labor unions. In these programs one is more likely to find the community- based educators who subscribe to a learner-centered, participatory method of program development and who engage the workers in the identification of their needs and the design, development and delivery of programs.

There are, of course, workplace literacy programs that include management, labor and educator partnerships. The resulting workplace literacy programs are likely to reflect the interactive nature of the "top-down & bottom-up" processes and strive to meet the needs of both employers and employees. These types of programs seem to flourish when the educator member of the partnership is committed to meeting the needs of both the employer and the employees.

Knowledge Resources for the Design and Delivery of Workplace Literacy Programs

A number of resources are available for those interested in learning how to design and deliver job-linked literacy programs, or for managers with training personnel whom they would like to receive training in how to develop job-linked programs. Careful study of the following books and reports will provide a knowledge base for designing and delivering job-linked literacy programs.

Basic Skills for the Workplace (Taylor, Lewe, & Draper, 1991). This 514 page volume includes four major parts with seven chapters per part. Part 1, Understanding the Need for Workplace Literacy, includes chapters that discuss the history of workplace literacy, the need for partnerships in developing workplace literacy programs, understanding that there are no "quick fixes" for employee skills training, and issues in developing a proposal for funding workplace literacy projects. Part 2, Identifying Workplace Training Needs, includes chapters that deal with literacy task analysis, assessment of learner needs, and how to develop workplace literacy programs.

Part 3, Examples of Practice in Workplace Basic Skills illustrates programs with both "top-down" and "bottom-up" approaches to development. English-as-a-Second-Language programs are discussed along with examples of workplace basic skills programs. Finally, Part 4, Discovering Approaches for Program Development, provides bibliographic resources for developing workplace literacy programs, and two chapters discuss the issues and methods involved in evaluating workplace literacy programs. In the scope of its coverage, this is the most comprehensive volume available on the design, development, and delivery of workplace literacy programs.

*Readin',Writin', and 'Rithmetic One More Time: The Role of Remediation in Vocational and Job Training Programs (Grubb, et al, 1991).*This report reviews basic skills education in vocational and workplace literacy programs, expecially those funded under the Job Training Partnership Act. It categorizes programs into "skills and drills," "functional context," and "eclectic" including those that integrate basic skills and vocational skills and "whole language" programs that are of the "bottom-up" persuasion. The report is highly critical of the "skills and drills" approach, more tolerant, but yet cautious about the functional context approach, and most favorably disposed to the "whole language" approach. It is misleading somewhat about the functional context approach because it assumes that it must be job-related (p.86), but that is incorrect. Overall the report raises the

important issue of why so many adult basic skills programs are so ineffective and it advocates more research into effective programs.

*Evaluating National Workplace Literacy Programs (Sticht, 1991).*This report was prepared for the U. S. Department of Education's National Workplace Literacy Program (NWLP) which has now been discontinued as such, though the new Adult Literacy and Family Literacy Act of 1998 provides support for workplace literacy programs. It discusses evaluation of workplace literacy programs funded by the U. S. government in order to meet the criteria that the NWLP used in evaluating proposals for workplace literacy programs. This includes evaluating how well programs establish the need for the program, the various program factors (such as program site location, instruction, etc.), quality of training, plan of operation, experience and quality of instructional personnel, and the evaluation plan and analyses to establish the cost-effectiveness of the program.

*What Work Requires of Schools (The Secretary's Commission on Achieving Necessary Skills (SCANS), 1991.*This report draws distinctions between the knowledge and skill requirements for "low performance" and "high performance" workplaces. The former are "Tayoristic," they engineer the demand for cognitive skills out of work through the assembly line approach. The latter are governed by "total quality management-TQM" concepts and they engineer cognitive skills back into the workplace by empowering workers to take charge of their products, work schedules, customer relations, and quality control. The SCANS report identifies five areas of competence and a set of foundation skills that it says should be taught to all school children and all employees so that America can compete more effectively for higher value added jobs in the world marketplace. The SCANS competencies and foundation skills provide resources for workplace literacy providers to incorporate into their designs for job-linked literacy programs. National efforts are underway to create certificates of mastery that will certify high school students and workers as competent in the SCANS competencies and foundation skills.

Workplace Basics (Carnevale, Gainer, & Meltzer, 1990). The American Society for Training and Development (ASTD) conducted a thirty-month research project to identify training practices in American businesses and industries. The study focussed on the skills that employers wanted employees to possess. This book reports on the results of the project's examination of the basic skills that corporations want in their workforce. It sixteen chapters it identifies and discusses the teaching of sixteen workplace skills: learning to learn, reading, writing, computation, oral communication, listening, problem solving, creative thinking, self esteem, motivation/goal setting, employability/career development, interpersonal skills, teamwork, negotiation, organizational effectiveness and leadership. The last chapter in the book provides guidelines for establishing effective workplace basics programs. Overall, the book reflects a more "top-down" point of view. At times it makes recommendations that are not sound. For instance, three pages are devoted in the chapter on Teamwork to the use of the Myers-Briggs Type Indicator which attempts to identify various personality types and how they would react in teamwork. However, the Myers-Briggs Type Indicator (MBTI) was recently evaluated for the U. S. Army Research Institute for the Behavioral and Social Sciences by the National Research Council of the National Academy of Science (Druckman & Bjork, 1991). That group concluded that "At this time, there is not sufficient, well-designed research to justify the use of the MBTI in career counseling programs." (p. 101) Other discussions of practices based on esoteric psychological ideas (The Johari Window; Jungian Theory of Personality Types) should be regarded with skepticism.

The Complete Theory-to-Practice Handbook of Adult Literacy (Soifer, et al, 1990). This is an easy to read exposition of the philosophical approach to literacy instruction known as "whole language." This approach is learner-centered and includes the participation of

learners (employees in the case of job-linked literacy programs) in the definition of skill and knowledge needs, the development of programs, the conduct of learning experiences, and the evaluation of programs. The book first describes the whole language framework. Then it discusses the development of programs in reading, writing, arithimetic, General Educational Development (GED) preparation, and the use of computers in adult literacy programs. There is a chapter that discusses staff selection and development that emphasizes the importance of the philosophical point of view that people hold regarding adult learning and adult learners in their employment as teachers. Finally, the last chapter discusses activities involved in program development and management. In the course of the book's treatment of instructional development, it discusses such phenomena as "invented spellings" in writing, "active reading" strategies including activities to do before, during, and after reading, and alternatives to standardized tests for assessing growth in learning, including portfolios of completed activities, vocabulary word banks, and other performance-based products for the assessment of learning. The book is a good presentation of the "bottom-up" approach to workplace literacy and is a favorite with many labor union educators.

Basic Skills Training: A Launchpad for Success in the Workplace; Literacy Task Anslysis: A How to Manual (Taylor & Lewe, 1990). The first report provides a brief overview of workplace literacy concerns and then gives examples of the results of literacy task analyses for several jobs: Motor Vehicle Repair; Grocery Store Receiver, Construction, etc. The second report is a step-by-step guide to conducting a literacy task analysis. Both are associated with "top-down" approaches.

Worker-Centered Learning: A Union Guide to Workplace Literacy (Sarmiento & Kay, 1990). This report represents the "bottom-up" approach to workplace literacy program development. It discusses the limitations of literacy task analysis for identifying the educational needs and desires of employees, focussing on the narrowness of such training and its limited generalizability. It challenges the "top-down" approach of "blaming the worker" for the loss of competitiveness, and advocates organizational change to "high performance" industries, and the active involvement of workers in the design and delivery of workplace literacy programs that can develop employee skills to the high levels needed by such organizations.

How to Gather and Develop Job Specific Literacy Materials for Basic Skills Instruction (Drew, Mikulecky, & Pershing, 1988). This practitioner's guide is also representative of "top-down" approaches. It discusses such topics as, Why take a job-literacy approach?, What is literacy task analysis, Needs Analysis, and Conducting literacy task analysis. The appendices provide sample lessons for job-related literacy training and provides instructions for using readability formulas to evaluate the reading skill levels of materials used at work.

*The Bottom Line: Basic Skills in the Workplace (U. S. Departments of Labor & Education, 1988).*This report marked the first official guidance of the U. S. government for how to do workplace literacy programs. It discusses the need for basic skills training in the workplace, how to identify workplace literacy problems (including how to conduct a literacy audit), designing, delivering, and evaluating workplace literacy programs. It includes questions that can be asked by a business in choosing a literacy provider. It is employer-oriented in a "top-down" manner.

*Cast-Off Youth: Policy and Training Methods from the Military Experience (Sticht, Armstrong, Hickey, & Caylor, 1987).*This book summarizes the history of functional context education and training research in the U. S. military services. It reviews seven research projects that illustrated the effectiveness of functional context methods in the

design and delivery of more effective technical training programs (electricity & electronics; medical; radio operators; etc.) and the design and delivery of literacy programs that integrate job-related materials into the curriculum. It provides a review of concepts from the cognitive sciences that are relevant to understanding the nature of literacy and its development from childhood to adulthood. It illustrates how the cognitive science concepts have been used to develop job-linked literacy programs, and it describes the development of a prototype electronics technicians' program that fully integrates technical knowledge and basic skills (reading, mathematics, problem solving) into the technical program. It summarizes the principles of functional context education and how the application of these principles to academic, technical, or basic skills education can facilitate entry into the learning environment, learning during the program, and transfer beyond the program.

Job-Related Basic Skills: Cases and Conclusions (Sticht & Mikulecky, 1984). This report provides a discussion of basic skills problems in the workplace in the early 1980's and gives three examples of job-linked literacy programs. The final section provides guidelines for developing occupationally related basic skills programs, including the need to have a conceptual framework for adult basic skills development, understanding principles of instructional systems development, the importance of time on task and the need for sound evaluation of programs.

Using the Knowledge Resources to Learn How to Design and Develop Job-Linked Literacy Programs

An employer may want people on his or her staff to acquire competence in designing and delivering workplace literacy programs. If so, and if there are already training and education professionals on the company staff, they could be assigned to develop this competence. They would presumably already possess a large background of knowledge about instructional design, educational methods, and how to evaluate the products of vendors who may offer workplace literacy programs and materials.

If you, as a staff member of a company wish to learn to design and deliver job-linked literacy programs it would be useful to first acquire the above resources. Then, follow an *active reading strategy* in which you do something before, during and after reading each book or report to get your background knowledge about the topic mobilized and to relate what you are learning to what you already know. This may involve previewing the tables of contents and trying to predict (guess) what each chapter will be about, taking notes during reading, or transforming the information from text to checklists or outlines, or summarizing what you have read, and then going back and to review what you have read. Start with the report on The Bottom Line. Read it and develop an understanding of the general steps it recommends for developing workplace literacy programs.

If you are a trainer or human resources development professional in a corporation you may wish to take *The Bottom Line* in hand and try to conduct a very cursory audit of the literacy materials and tasks that the employees you are concerned about must face. If you are planning major technological or organizational structural changes, e.g., introducing TQM, then try to anticipate the new demands for skills that such changes might engender. Read the SCANS report.

After that, locate and visit two or three workplace literacy programs in your vicinity. Have the staff and students explain what they are trying to accomplish. Look at the kinds of materials they are using. Find out how they determined what and how to teach the program and how to determine if the program is meeting their goals for it.

Then go back to the resource materials. Turn to the first resource book given above (Basic Skills for the Workplace) and select chapters that provide an overview of workplace literacy concepts, then search for examples of such programs, and then read chapters dealing with task analysis, program development, and evaluation. Compare what you are reading to what you observed and talked about at the programs you visited.

Next read the book on *The Complete Theory-to-Practice Handbook of Adult Literacy. Contrast* the approach it discusses with that of *The Bottom Line.* Notice what is meant by the "whole language" approach; what is meant by learner-centered, participatory methods of program design; what the "language experience" method of teaching language and literacy skills includes; how computers can be used to teach writing and analysis processes, such as those involved in using word processing programs and spreadsheets.

Read the report on *Worker-Centered Learning* and compare the procedures for program development there with those discussed in the preceding volume and in *The Bottom Line* and the *Basic Skills for the Workforce.* Distinguish between the concepts of "top-down" and "bottom-up" approaches developed in this paper.

Visit a program developed in a "bottom-up" approach. See how curriculum materials are developed, how instruction is carried out, and how evaluation is accomplished. Develop a list of "pros" and "cons" for each approach. This will be helped by reading *Readin, Writin,' and 'Rithmetic* and noting the differences between the "skills and drills," "functional context," and "eclectic" programs (including the "whole language" programs) critiqued in the report.

It may be useful to think about whether you are interested in designing a vestibule workplace literacy program for hiring under skilled persons and then raising their skills to meet entry level job requirements, or whether you are interested in programs for retraining employees in the face of organizational or technological change (or both), or for upgrading employees skills for promotability, or for retraining employees for outplacement in the course of a reduction in force.

Vestibule programs are more likely to involve "top-down" curriculum development methods because the new employees are not familiar with the jobs and what they must learn to qualify for full entrance into a job. The other changes that may instigate the need for a job-linked literacy program can rely more on 'bottom-up' curriculum development methods that place more responsibility on employees for thinking about their present and future educational needs.

Read the remaining resource materials to develop an understanding of the various issues involved in teaching reading, writing, arithmetic, and various other cognitive skills in the context of job- or work-related knowledge. Also, it would be well to recall that education may also be helpful to employers and employees if it helps employees adapt better off the job and in the community. Therefore, job-linked programs may be considered in turns of the degree to which they focus directly on job-tasks, the work culture and setting, or are less directly related, but nevertheless important to work performance, and involve basic skills use in community, home and school settings.

In some cases job-linked literacy programs may be viewed as the first step toward getting employees engaged in lifelong learning activities. This may result from the necessity to constantly learn new job-related tasks, or to achieve new levels of competence if "pay for competence" methods begin to replace time on the job (seniority) as the basis for promotions and pay raises.

The "proof of the pudding" as to how well the concepts of job-linked literacy program design and delivery have been learned will come with the actual design and delivery of a program. Inevitably, the first attempt will suggest adaptations for a new iteration. Further study of the knowledge resources given above, and new resources that are encountered will contribute to a growing competence on the part of corporations, unions, and educational providers in meeting the needs for job-linked literacy programs.

References

Carnevale, A., Gainer, L. & Meltzer, A. (1990). *Workplace Basics*. San Francisco, CA: Jossey-Bass.

Drew, R., Mikulecky, L., & Pershing, J. (Eds.). (1988). *How to Gather and Develop Job Specific Literacy Materials for Basic Skills Instruction. A Practitioner's Guide.* Bloomington, IN: School of Education, Indiana University.

Druckman, D. & Bjork, R. (Eds.)(1991). *In the Mind's Eye: Enhancing Human Performance.* Washington, DC: National Academy Press.

Freire, P. (1970). *Pedagogy of the Oppressed.* New York: Seabury Press.

Grubb, W., Kalman, J., Castellano, M., Brown, C., & Bradby, D. (1991, September). *Readin',Writin', and 'Rithmetic One More Time: The Role of Remediation in Vocational Education and Job Training Programs.* Report No. MDS-309. Berkeley, CA: National Center for Research in Vocational Education.

Johnston, JW. & Packer, A. (1987). *Workforce 2000: Work and Workers for the Twenty-First Century.* Indianapolis, IN: Hudson Institute.

Philippi, J. (1991). *Literacy at Work: The Workbook for Program Developers.* New York: Simon & Schuster, Inc.

Sarmiento, A. & Kay, A. (1990). *Worker-Centered Learning: A Union Guide to Workplace Literacy.* Washington, DC: AFL-CIO Human Resources Development Institute.

Soifer, R., Irwin, M., Crumrine, B., Honzaki, E., Simmons, B., & Young, D. (1990). *The Complete Theory-to-Practice Handbook of Adult Literacy.* New York: Teachers College Press.

Sticht, T. (1991, April). *Evaluating National Workplace Literacy Programs.* Washington, DC: U.S. Department of Education, Adult Literacy Clearinghouse.

Sticht, T. (1990, January). *Testing and Assessment in Adult Basic Education and English as a Second Language Programs.* Washington, DC: U.S. Department of Education.

Sticht, T. & Mikulecky, L. 1984). *Job-Related Basic Skills: Cases and Conclusions.* Columbus, OH: ERIC Clearinghouse on Adult, Career, and Vocational Education.

Sticht, T., Armstrong, W., Hickey, D., & Caylor, J. (1987). *Cast-Off Youth: Policy and Practice from the Military Experience.* New York: Praeger.

Taylor, M. C. & Lewe, G. R. (1990, December). *Basic Skills Training: A Launchpad for Success in the Workplace.* Ottawa, Ontario, Canada: Algonquin College. (613) 564-5439)

Taylor, M.C. & Lewe, G.R. (1990, December) *Literacy Task Analysis: A How To Manual for Workplace Literacy Trainers.* Ottawa, Ontario, Canada: Algonquin College. (613) 564-5439

Taylor, M.C., Lewe, G. R., & Draper, J. A. (Eds.). (1991). *Basic Skills for the Workplace..* Toronto, Ontario, Canada: Culture Concepts, Inc.

U.S. Departments of Education and Labor. (1988) *The Bottom Line: Basic Skills in the Workplace..* Washington, DC: U.S. Department of Labor.

Secretary's Commission on Achieving Necessary Skills. (1991). *What Work Requires of Schools.* Washington, DC: U.S. Department of Labor.

Chapter 2

Q & A on the Evaluation of Workplace Literacy Programs

This chapter is based on a paper prepared for the National Workplace Literacy Program (NWLP) of the U. S. Department of Education (Sticht, 1991), which has since been replaced by the Workforce Investment Act, Title 2: Adult Literacy and Family Literacy Act of 1998, and the Work in America Institute's Job-Linked Literacy Network (Sticht, 1992). The purpose of the latter paper was to stimulate discussion of issues related to the evaluation of workplace (job-linked) literacy programs. In preparing the paper, a Question & Answer (Q & A) format was followed and responses were prepared to four questions asked by the Work in America Institute. Those same questions are presented here in bold type, followed by their responses, which have been enlarged to include part of the NWLP paper, too. No claim is made that responses are complete and nor that they fully explicate all the nuances and chains of thought that the various questions stimulate. And they certainly do not form the last word on what could be said about these complex issues.

By what criteria should a company judge the value of its program?

Clearly, a company should use criteria for evaluating its program that reflect the goals of the program. That is, the company needs to know whether the program is accomplishing the goals that it has set for the program. This, then, produces the problem of how to establish goals for the program. If one knows what one wants to achieve with the program then criteria that reflect those achievements can be developed.

For example, if one goal of a workplace literacy program is to improve people's ability to read job-related materials, then it is a reasonable question to ask, "Can people who have completed the workplace reading program read their job-related materials better after completing the program than they could before they took the program?"

In turn, this raises the important questions of how many program participants should improve their abilities to read job-related materials to what extent? If 100 employees take a job-related reading course that teaches 100 applications of reading as used on the job, should all who take the course master all job-related reading tasks? And should they all be expected to do this within the same period of time? For instance, in the period of one 45 hour program.

A Case Study With Evaluation Data. In one hospital-based workplace literacy program (Nurss, 1990), pre-and post-program tests were constructed based on information in the employee handbook and job memos that were applicable to all departments. The reading tests were 20 cloze tests (cloze tests are constructed by deleting every fifth word in a passage. The examinee then guesses what the missing word is. This type of test is highly correlated with other types of reading tests).

In the hospital program, the average pre-test score was 69% and the post-test score was 74%, and the difference of 5 percentage points was statistically reliable. In this program, participants clearly did not reach 100 percent mastery. However, in assessments using cloze tests, people hardly ever reach 100 percent mastery. The question is, is the 5 percentage points (about 1.03 raw score points) increase indicative of important and sustained improvements in job-related reading? In this program evaluation, participants also made statistically reliable improvements in written and oral communication - in about the same percentage range of improvement.

Employees also were interviewed and reported their perceptions of the programs effects: 61% reported that participation in the program increased their academic (reading) skills, 39% improved their oral skills, 34% their written expression, 29% their job knowledge, 27% improved their self-confidence, and 24% reported that they improved their basic education. Many participants (69%) reported that their reason for attending the classes was to get a promotion or a better job. At the end of the program two (3%) had achieved this goal.

In this program the hospital staff, education providers, and external evaluator evaluated the program as successful, citing employee improvements as the basic criteria for evaluating the program as successful.

A Policy-Oriented Study With Minimal Data. A report by the Southport Institute for Policy Analysis (Faison, Vencill, & McVey (1992) describes how four small manufacturing firms provide "basic skills" or "workforce literacy" instruction to their employees. The report was accompanied by a press release stating that the study "...shows that both management and workers are enthusiastic about the results."

Among the benefits cited by managers in the four firms were "...less waste of time and materials, reduced error rates, greater customer satisfaction, improved communications and labor relations, and the ability to introduce new production processes and systems of work organization." The press release also quotes two workers, one stating that, in regard to a math course taken, "I have more confidence; I can do in a half hour what used to take two hours." Another stated, "I can do reports faster because my English is better. I can say what I need to say and show the boss I'm interested in doing the job and getting ahead."

Significantly, this report, by a policy analysis institute, includes no quantitative data showing pre- and post-test scores such as those in the hospital case study cited above. There are, in fact, no data on percentages of employees who reported reaching their goals, and, in fact, no clearly stated goals are given for any of the four firms studied. If each of the four firms had as one of it's goals that participating employees should improve their ability to read their job-related reading materials, from this study we would not be able to say how many achieved this goal.

Limitations of Self Report Data. The most frequently occurring information in reports of the outcomes of job-linked literacy programs are the perceptions of program participants, instructors, and supervisory or management staff. While such information can be informative, it must be regarded with caution. In one study (Sticht, 1975) teachers in job-related reading programs at four locations in the nation were asked to estimate gains in pre- to post-test scores of job-related and 'general" reading. At all four sites, teachers tended to overestimate general reading gains by from 1 to 2 "years" over what test score data indicated. At three of the four sites, teachers overestimated job-related reading gains by 1 to 2.5 "years" and underestimated gains by several "months" at the fourth site. These data suggest that, even when teachers have access to data, they frequently misjudge program effects on reading skills.

In another study (Bonner, 1978), 108 army infantrymen were asked to assess their ability to perform six important job tasks to the standards required for their work. Sixty-eight supervisors were then asked to rate the abilities of these infantrymen to perform the six tasks. Finally, the infantrymen were asked to actually perform the tasks in hands-on job sample tests. The results showed that on five of the six tasks, both the infantrymen and their supervisors greatly overestimated the ability of the personnel. For instance, on installing and recovering an electrically armed claymore mine, 64 percent of the infantrymen said they could do the task. Fifty-one percent of their supervisors said the infantrymen could do the

task. But on the job sample task, only 8 percent could actually perform the task to standards!

On average, 63 percent of infantrymen and 56 percent of their supervisors said they could perform the six tasks, while the hands-on job sample tests showed that only 37 percent of the tasks could be performed to standards.

The findings of studies of teacher's, supervisor's, and employee's judgments cited above (and others not cited here) indicate that caution should be placed on evaluations of the effects of basic skills programs on basic skills improvement and job performance that are based on such judgments. If it is difficult for employees and their supervisors to accurately judge their actual job ability, it may be even more difficult for them to accurately judge how literacy training affects their job performance.

In cases wherein supervisors say, for instance, that wastage is down, or there is greater customer satisfaction, as in the work of the Southport Institute discussed above, it should be possible to quantitatively document the pre- to post-program rates of wastage and customer dissatisfaction and to then describe the probability paths between the job-linked literacy program and the reduction in wastage, improved customer relations, and so forth.

It is ironic that such quantitative documentation is called for in the total quality management (TQM) philosophy that has lead so many companies to initiate workplace literacy programs. Yet the same emphasis upon statistical process control that is the foundation for implementing TQM (e.g. in designing quality in at each step in a production line, not inspecting it in at the end of a production line) is not being applied to any great extent to the job-linked literacy programs that are expected to promote the effectiveness of TQM.

How should the government evaluate the programs it funds?

Like business and industry, government agencies should evaluate workplace literacy programs according to the goals that the government (representing the general population of the nation) has for such programs.

Typically, whenever the federal government becomes involved in funding educational programs, there is a need for the government officials to review programs to determine whether the programs they have funded are, in fact, providing useful educational experiences that meet the intent of the Congress, as representatives of the public at large. In this case then, it is advantageous to go beyond the self-reports of those involved that they are receiving beneficial educational services. There is a need for additional evidence of the effectiveness of the program that is less subjective. For instance, if a program aims to improve the ability of employees to read their job-related materials, then it is not sufficient for evaluation to report that instructors and employees *say* they can read their job-related materials better after they have been in the program for a while. Rather, some confirming evidence, such as demonstrated improvements in performing job-related reading tasks, would be useful.

For the federally funded National Workplace Literacy Programs of the late 1980's and mid-1990's, the U. S. Department of Education published rules and regulations regarding the evaluation of such programs (The Federal Register, Friday, August 18, 1989, pp. 34418-34422). Among other things, the regulations required that each application for funds under the program include an evaluation plan (see Table 2.1, column 6). In this case then, program operators had to satisfy not only themselves and the other participants active in the

program as to the value of the program, they also had to satisfy the Department of Education which was required to report on the value of the programs to Congress.

Good evaluation starts at the beginning, not the end of a program. For this reason, the discussion of evaluation below focuses upon the relationship of the evaluation of workplace programs to the original criteria that the Secretary of Education developed to evaluate proposals applying for funds to establish programs. Table 2.1 presents the criteria used to evaluate proposals, reworded and rearranged here to emphasize their use in preparing an evaluation report of a program once it has been funded and implemented. These criteria specify, in broad outline, what a well-designed and operated workplace literacy program would look like. The evaluation, then, indicates (1) how well the program operators implemented the design and operational plans that they submitted for funding, (2) what outcomes are being achieved and (3) how the program can be modified to make it more effective.

In general, the purpose of evaluation for the Department of Education is to permit the Department to place a value on a given program in providing services and in demonstrating innovative and effective practices. That is, it must first decide whether a proposal for a program is likely to result in a needed and effective program, and then it must decide whether the program finally developed and implemented provided an educational experience that met the stated criteria outlined in the original proposal and the intent of the Congress when it passed legislation funding adult literacy education.

Evaluation is not something to be accomplished at the end of a program development and implementation effort to "see if it worked." Rather, evaluation is an integral part of the original design of the program and an ongoing process that can permit decisions about how well the program is achieving one or more of the purposes of the Adult Education and Family Literacy Act and, where desirable, to improve the program and its value to adult learners, other partners in the project, and the society at large.

Purpose of the National Workplace Literacy Program. When the NWLP was in operation, both the literacy providers and the Department of Education had to evaluate their programs with regard to how well they were achieving the purpose of the National Workplace Literacy Program (NWLP). Figure 2.1 outlines the general purpose of the NWLP and illustrates the types of literacy and productivity indicators that might be included in a workplace literacy program at present.

The general purpose of the NWLP was to provide grants or cooperative agreements involving exemplary partnerships of business, industry, or labor organizations and educational organizations for projects designed to *improve the productivity of the workforce* through the *improvement of literacy skills in the workplace* by -

(a) Providing adult literacy and other basic skills services and activities;

(b) Providing adult secondary education services and activities that could lead to the completion of a high school diploma or its equivalent;

(c) Meeting the literacy needs of adults with limited English proficiency;

(d) Upgrading or updating basic skills of adult workers in accordance with changes in workplace requirements, technology, products, or processes;

(e) Improving the competency of adult workers in speaking, listening, reasoning, and problem solving; or

(f) Providing educational counseling, transportation, and child care services for adult workers during non working hours while the workers participate in the project (Federal Register, August 18, 1989, vol. 54, no. 159, p. 34418).

The NWLP aimed to improve the productivity of the workforce by improving the literacy of the workforce. This leads to the two primary questions for evaluation: (1) does the program improve workforce literacy abilities, and (2) do the improved literacy abilities lead to improved productivity?

The Relationship of Literacy Ability to Productivity. The basic assumption of the NWLP was that there is a relationship between various literacy abilities and job productivity, as indicated by various measures.

Though this may seem straightforward, it is not true that all aspects of productivity are directly mediated by literacy ability. For instance, many job tasks do not require the direct application of reading or writing abilities. Nor will they necessarily require specialized knowledge that requires reading and writing abilities. Many job tasks can be learned by watching others and imitating them.

Therefore, in determining the need for a workplace literacy program that emphasizes increasing the reading, writing, or other literacy abilities of the workforce, it is important that program developers understand the role of literacy ability in relation to various indicators of productivity. Otherwise, if there is simply a blanket assumption that increasing literacy ability will increase productivity in some unspecified manner, it may not be possible to demonstrate that the program has, indeed, increased productivity.

Some productivity indicators may be directly mediated by literacy abilities while others may be only indirectly mediated by literacy ability. For example, being able to comprehend oral directions that supervisors provide is directly mediated by the ability to comprehend the English language, if that is the language used by the supervisor. If the directions are not understood, then the worker may not know what to do or how to do it. In this case, the job tasks may not get done, or they may not be correctly performed, even though the tasks, themselves, do not require language comprehension.

In such circumstances, improving English language comprehension skills may lead to improved job task performance not because the tasks require language comprehension, but because understanding the directions about what to do and how to do it requires language comprehension.

On the other hand, because the job tasks do not directly involve the comprehension of English language, it is possible that workers may learn what to do and how to do it by watching others. In this case, then, increasing English language skills may not lead to improved task performance. Therefore, some other indicator of the increase in productivity due to increased language ability should be sought.

Generally speaking, *unless a direct relationship to some indicator of productivity can be demonstrated in the design of the program, the program developer should not promise to improve that aspect of job productivity.* However, as a part of the program evaluation, information about aspects of productivity that are not known to be directly mediated by literacy ability should be obtained because of the possibility of the indirect influence of increased literacy ability, or simply participation in the literacy program, may have on various indicators of productivity. For instance, if having access to education programs

boosts employee morale, indicators of productivity such as attendance, less tardiness, increased co-operation (team work) and so forth may improve.

Table 2.1. Illustration of How Criteria for Evaluating Proposals for National Workplace Literacy Programs Can be Used to Report Evaluations of Programs.

Need for the Project 1	Program Factors 2	Quality of Training 3
Documents the needs to be addressed by the project.	Demonstrates the active commitment of all partners to accomplishing project goals.	Provides training through an educational agency rather than a business, unless transferring training to a business is necessary and reasonable within the framework of the project.
Focuses on demonstrated needs of adults for workplace literacy training	Targets adults with inadequate skills aimed at new employment, career advancement, or increased productivity.	Delivers instruction in a readily accessible environment conducive to adult learning.
Documents how needs will be met.	Includes support services based on cooperative partnerships to overcome barriers to participation by adult workers.	Uses individualized educational plans developed jointly by instructors and adult learners.
Documents benefits to adult workers and their industries that will result from meeting those needs.	Demonstrates a strong relationship between the skills taught and the literacy requirements of actual jobs.	Uses curriculum materials designed for adults that reflect the needs of the workplace.

Plan of Operation 4	Experience & Quality of Personnel 5	Evaluation Plan & Cost-Effectiveness 6
Describes roles of each member and each site of the partnership.	Provides evidence of the applicant's experience in providing literacy services to working adults.	Provides clear, appropriate methods of evaluation that are objective and produces data that are quantifiable.
Describes activities to be carried out by any contractors.	Provides evidence of the experience and training of the project director in project management.	Identifies expected outcomes of the participants and how those outcomes will be measured.
Describes roles of other organizations in providing cash, in-kind assistance, or other contributions to the project.	Provides evidence of the experience and training of key personnel in relation to the project requirements.	Determines effects of program on job retention, performance, and advancement.
Describes the objectives of the project and plan to use project resources to achieve each objective.	Indicates amount of time each key person will devote to project.	Obtains data that can be used for program improvement.
Establishes measurable objectives for the project that are based on the project's overall goals.	Indicates how nondiscriminatory employment practices will be implemented.	Provides data indicating costs of the program in relation to its benefits.

Source: Federal Register, Vol. 54, No. 159, Friday, August 18, 1989, pp. 34419-34420. Note that the wording and ordering here is not the same as in the federal regulations. The latter should be used for exact wording. The present ordering is for illustrating how the criteria may be used for the evaluation of programs not proposals.

Relationship of Program Design and Development to Evaluation

Because the purpose of the NWLP was to increase workforce productivity through the improvement of literacy ability, the design of a workplace literacy program should have indicated the relationship between literacy ability and productivity, and how the program intended to increase productivity through the improvement of some aspect of literacy ability.

This relationship of program design and development to evaluation is illustrated in Table 2.1 in columns 1, 2, and 3. Column 1 calls for a *needs assessment* that focusses on documenting the needs of adults for workplace literacy training, how the needs will be met and how meeting those needs will benefit the workers and their industries. Column 2 calls for program factors that demonstrate a strong relationship between the skills taught and the literacy requirements of actual jobs. Then Column 3 makes clear the need to directly address the program to workplace literacy requirements by calling for the use of curriculum materials that reflect the needs of the workplace.

If the design of the program accomplishes the activities of columns 1,2, and 3, then the program will have gone a long way toward meeting the requirements of Column 6 for the identification of expected outcomes, how those outcomes will be measured, and how those outcomes are related to job retention, performance and advancement.

Using Table 2.1 in Program Evaluation. As illustrated above, Table 2.1 outlines criteria for a well-designed workplace literacy program. Presumably, since these are criteria used to select projects for funding, any projects that receive funding have successfully met these criteria, at least to some minimally acceptable extent.

The process of evaluation is the process of turning the various declarative statements, such as "Focuses on demonstrated needs of adults for workplace literacy training"(column 1) into questions, such as "Does the program focus on the needs of adults for workplace literacy training, and how is this demonstrated?" By following this procedure of transforming declarative into interrogative statements, Table 2.1 can be transformed from a list of criteria for evaluating *proposals* for programs into criteria for evaluating *programs* of workplace literacy.

Table 2.2 illustrates how the categories of Table 2.1 can be used to summarize the results of evaluation studies. In Table 2.2, findings are summarized from a study of the National Workplace Literacy Program by Kutner, Sherman, & Webb (1990; source number 1). Additionally, results are summarized from a survey of workplace literacy program evaluations by Mikulecky & D'Adamo-Weinstein (1990; source number 2). While the placement of the particular findings in the Table 2.1 categories may be arguable in some cases, the point is that Table 2.2 illustrates that the categories of Table 2.1 may be used to conduct and report evaluations of NWLP projects.

For instance, note that in Table 2.1, category 2-Program Factors, calls for the proposal to "Demonstrate a strong relationship between the skills taught and the literacy requirements of actual jobs." Then, in Table 2.2, category 2, it is noted that "Study sites typically assess participant literacy levels through standardized tests that are typically used for ABE and are not geared for workplace literacy." Because standardized tests do not strongly represent "the literacy requirements of actual jobs" they were not considered appropriate for assessing participant literacy levels. This observation was included in category 2, rather than in category 6-evaluation- because it illustrates the difficulty of matching skills taught (and assessed) to the literacy requirements of actual jobs.

Table 2.2. Comments from Evaluations of Workplace Literacy Programs.

Need for the Project 1	Program Factors 2	Quality of Training 3
Supervisors are involved with the workplace literacy projects at many of the business sites. Initial reluctance of supervisors at many of these sites to have workers attend classes on company time has been eliminated as benefits from the project have become apparent. (1)	Although business sites are supportive of the respective workplace literacy projects, few indicated a commitment to continue the project without either federal or other outside funding. (1) Study sites typically assess participant literacy levels through standardized tests not geared for workplace literacy. (1)	A number of project components may contribute to the absence of retention problems: locating instructional services at the work site, providing participants with monetary incentives, offering a supportive learning environment, support services, transportation, and counseling. (1)
Formal literacy task analyses is the exception rather than the rule.(1) Businesses at the study sites are actively involved with recruiting participants by identifying potential participants. (1)	Educational providers at the study sites are directly responsible for all instruction-related activities, including conducting literacy task analyses, assessing the literacy skills of participants, developing instructional materials, and hiring and managing instructors. (1)	There is substantial variation from site to site in the total number of hours available per training cycle.(1) When instructors do not share program goals and resources are inadequate instructional quality is likely to be inadequate. (2)

Plan of Operation 4	Experience & Quality of Personnel 5	Evaluation Plan & Cost Effectiveness 6
Business partners at the study sites are not heavily involved with the day-to-day activities of the workplace literacy projects. (1)	With only one exception, educational providers at the study sites do not have prior experience with workplace literacy. (1)	Study sites do not generally conduct formal evaluations of their projects. (1) Learners were often evaluated by supervisors in informal reports. (2)
Increased demands for classes are reported as indicators of program success. (2)	Almost all of the educational providers at the study sites have hired instructors who possess experience with ABE or ESL programs. (1)	Program evaluations tend to be informal with little or no empirical data. (2)
Anecdotal experiences are reported as indicators of program success. (2) Evaluations generally rely on ancedotal evidence, including the perceptions of instructors, business supervisors, and more senior staff. (1)	Most educational providers at the study sites do not provide training for instructors before instructional services began. Most, however, do offer in-service training for instructors and volunteers. (1)	When programs are evaluated, they are often assessed mainly through completion of questionaires and/or surveys by program participants. (2) Some programs do test participants both before and after completing the program. These results are often reported only in general terms as indicators of progam effectiveness. (2)

Sources: (1) Kutner, Sherman & Webb, 1990; (2) Mikulecky & D'Adamo-Weinstein, 1990.

Additional entries in Table 2.2 suggest the types of findings that professional evaluators have reported from their studies of workplace literacy programs. They illustrate, therefore, the kinds of activities and problems that others might consider in evaluating workplace literacy programs.

The Need for Data On Program Effectiveness

Perhaps the most vexing problem in program evaluation is the determination of whether the outcomes that are achieved are useful and justify the expenditures of public funds for this activity to meet learner needs rather than for something else. One of the reasons this is such a problem is that, while this type of decision making is necessary at the federal level, it is not the major concern of local workplace literacy programs. In these programs, program administrators and teachers are concerned with meeting the needs of their adult learners and partners. They are less concerned, if at all, with meeting the needs of federal funding agencies for information for decision making.

While obtaining convincing outcome data is difficult because it is not the highest priority for workplace literacy teachers and adult learners, the problem is compounded by the fact that hundreds of millions of dollars are spent each year on standardized tests and other assessment instruments, throughout the education system in the U.S., and yet no one is satisfied that they are actually obtaining valid information about "true" achievements. This is indicated by the fact that today there are several national activities underway to develop new national examinations to obtain a more valid indicator of how well the nation is doing in education.

In the face of such difficulties in satisfying ourselves that we are doing good, bad, or so-so with regard to educational achievement across the spectrum of educational services in the nation, it is understandable why workplace literacy operators, teachers and adult learners may be reluctant to submit to examinations that they feel are intrusive and non representative of what they are teaching and learning.

Nonetheless, the fact remains that there is a need, at the federal level, for information regarding the effectiveness of the learning activities and outcomes that are taking place under the Adult Education and Family Literacy Act of 1998. That is why the criteria for the earlier NWLP proposals includes column 6 of Table 2.1-Evaluation Plan & Cost-Effectiveness-which includes the requirements for methods of evaluation that are "objective" and which indicate how "outcomes will be measured."

These requirements for "objectivity" and "measurability" of outcomes in evaluation are not baseless requirements of the funding agency. As Table 2.2 indicates, outside evaluators who have examined workplace literacy programs have independently observed that "program evaluations tend to be informal (unstandardized) with little or no empirical (objective) data (quantifiable measures)".

In fact, the repeated findings by outside evaluators that programs lack "formal" evaluations, that they use "informal" reports, depend primarily upon self-report questionaires with no substantiating evidence in more "objective" terms of what is reported, and provide "little or no empirical data" are among the most salient outcomes of external evaluations of workplace literacy programs (and all other programs in adult literacy or ABE for that matter).

In short, what these evaluators say is needed is *convincing evidence* that useful learning outcomes are being achieved in adult workplace literacy programs, whether offered in the workplace or elsewhere, and that this new learning results in *improved productivity* in finding, retaining, performing, or advancing in a job in the workplace. While various types

of ratings (e.g., supervisor ratings of increased productivity; teacher ratings of improvement; adult learner ratings of pre-and post-program increases in learning or productivity) provide useful indicators of the program's effects on learning and productivity, such ratings are not totally convincing. They are not free of the potential for self-deception that may bias ratings.

It is the desire to overcome these kinds of subjective judgments that may lead to inaccurate or invalid estimates of the outcomes of programs that lead the federal criteria and evaluation experts to call for "objective","empirical", "measurable" outcomes of literacy learning and productivity.

Measuring the Outcomes of Learning. The goal of workforce literacy is to improve the literacy skills of the workforce and thereby increase workplace productivity (see Figure 2.1). Therefore, the primary outcome of a workplace literacy program that needs to be measured is the extent to which literacy abilities (defined broadly as the set in Figure 2.1) have been improved. However, it should be noted that some indicators of productivity may increase due to increased morale when a company shows employees that it cares enough to provide them an educational opportunity. Thus, a workplace literacy program may have an effect on productivity even when there is little or no measurable improvement in literacy abilities.

The measurement of literacy abilities ought to reflect the content of what is being taught. The latter, in turn, will have the best chance of being transferred to the job if it consists of the materials and content knowledge needed for getting and performing a job. For instance, if workers in a plant need to learn to write reports from production team meetings, it would be better to teach writing using the writing of team production reports as the vehicle for teaching proper usage of punctuation, planning, presenting, and revising a composition, and other aspects of English language, than to use the writing of fiction or personal accounts of one's life events.

The only way to know if growth has taken place in literacy abilities is to measure the abilities at the outset of the program, and then again later on. Typically, it will be possible to measure both the content knowledge that worker's have relevant to some new domain of learning that they wish to command, and the types of knowledge and skill that they possess regarding the uses of language and literacy in working with knowledge for doing something or learning something. For instance, developing job-related reading task tests (JRTT) using the materials from literacy task analyses (The Bottom Line, 1988) can permit the assessment of how much of the content knowledge in some job or work-related domain the worker knows and how well the worker can apply information search, comprehension strategies, and study skills to locate and learn knowledge that is not known. Administering JRTT as pre-and post-tests will permit an assessment of how much improvement has occurred in workplace reading skills.

While JRTT can indicate something of the growth of job-related literacy abilities, they do not permit comparisons of growth in one program with growth in another program by other workers. Yet the Department of Education needs to know how well programs perform relative to one another. For this reason, it is necessary to use one or another nationally normed, standardized literacy tests as pre- and post-program measures of the generalizability of growth (see Chapter 4 for an extended discussion of standardized testing in the context ot the Department of Education's adult basic education program).

In using such tests, care should be taken to not over estimate the growth that has taken place. This may happen if very large increases in test performance are obtained. For instance, if a worker makes a two to five year improvement in test scores in a 20 to 100

hour program, the gain should be suspected as inflated due to faulty testing circumstances at the pre-test, post-test, or both. For this reason, frequency distributions of pre- and post-test scores should be reported, not simply means or medians. The latter conceal the variability in the gain scores that evaluators can use to judge the extent to which testing artifacts may be influencing test performance.

Estimating the practical value of test score gains. As illustrated by the hospital case discussed above, it is not too difficult to obtain some indication of learning, as in the use of the cloze tests in the hospital program. In the hospital program the pre- test score on the cloze test was 13.79 while the post-test score was 14.82, an improvement of 1.03 raw score points. But is this practically useful? The problem from the federal government's perspective is to understand just what a five percent gain in cloze test scores means in terms of gain in the employees' literacy skills. Is this a practically useful (not just statistically reliable) gain in skill? How does this improvement compare to improvements made in other programs? Could other approaches result in making more gain than the current program makes?

One methodology for *estimating* the practical usefulness of the differences between mean scores is to calculate the "effect size." The effect size is a percentage of the standard deviation of the pre-test. For instance, in the hospital case, the standard deviation for the pre-test cloze score distribution was 4.67. The mean gain of 1.03 points is 1.03/4.67= .22 percent of the standard deviation of the pre-test distribution of scores.

The meaning of the effect size is that, if the scores on the pretest are normally distributed, then the mean score of the pre-test group is at the 50th percentile. Then an effect size of .22 means that the mean score on the post-test is at roughly the 60th percentile of the pre-test distribution. Overall, then, it would be argued that the group had improved from the 50th to the 60th percentile, and, practically speaking, people at the 60th percentile tend to perform better on literacy tasks than people at the 50th percentile.

While the effect size methodology for comparing program gains can be useful, it generally requires a large number of participants and careful test development to ensure that test score distributions are normally distributed. In the case of cloze tests, if one changes the algorithm for constructing the test from the deletion of every fifth word counting from the first word, to the deletion of every fifth word counting from the second, or third, etc. word, then large changes may be obtained in the test scores. In one study, changes in the deletion algorithm resulted in raw score changes from 17 to 40, a 135 percent improvement in performance (Sticht, 1975, June, p. 25). This suggests the need for confirmatory data on growth in learning using additional methods of assessment.

Because of the need for comparative data on programs, the federal government generally suggests that workplace literacy programs use one of the several nationally-normed, standardized literacy tests to measure growth in job-linked literacy programs. This information can then be supplemented with performance data on locally developed indicators of achievement such as job-related reading task tests or cloze tests and the combined information can be used in reaching judgments about the beneficial effects of the program on learning.

Measuring Improvements in Productivity. A major goal of federally funded workplace literacy programs is to improve the productivity of the workforce through the improvement of worker's literacy abilities. For this reason, after providing convincing evidence that improvements have taken place in literacy abilities, the workplace literacy provider needs to present convincing evidence that the improvements in literacy have lead to improvements in job productivity. If the materials and tasks used in the literacy program are direct simulations

of tasks involving the use of literacy abilities on the job, then the JRTT or other literacy assessments *are* direct indicators of increased productivity in performing the literacy-mediated components of job tasks.

However, it is important to distinguish those aspects of productivity that can be shown to be directly mediated or affected by literacy abilities and those that are capable of being affected by factors other than increases in literacy abilities. *Workplace literacy programs should only be held accountable for improving those aspects of productivity directly mediated by literacy abilities.* And even then care should be exercised in building expectations for the effects of literacy education on productivity. Too many other factors, such as poor supervision, bad management practices, substance abuse, and so forth may influence productivity to expect improved literacy to overcome any and all productivity problems. Workplace literacy providers should not promise more than they can be certain of delivering when it comes to improving productivity.

One of the most frequently used methods of evaluating changes in productivity is to have supervisors provide pre-and post-program ratings of improvements in such factors as attendance, lateness for work, accuracy in performing job tasks, reductions in errors or wastage of material, compliance with safety rules, or other types of indicators of productivity (cf., Mikulecky & Lloyd, 1993). While this information is useful in evaluating the effects of literacy education on productivity, it is subject to the criticisms of subjective ratings given above. In this regard, it is useful to have ratings of literacy program participants and non-participants from supervisors who do not know which employees have been involved in literacy training. This reduces the likelihood of positive bias for program participants on the part of the supervisors.

If possible, company records of performance appraisals of participants before the literacy training and after should be obtained and summarized. Records of waste, returned products, customer complaints and other objective indicators of productivity should be sought to support the rating information. Additional examples of productivity measures can be found in the list of resources included with this paper.

Are current government requirements for evaluation realistic and useful for companies receiving government funds?

The criteria outlined in Table 2.1 for evaluating *proposals* for workplace literacy programs may be generally useful categories for evaluating workplace literacy *programs*, too. In some cases, however, some of the details of the categories may need modification. For instance, under category 2 - Program Factors - one of the desired characteristics of a workplace literacy proposal is that the program "Demonstrates a strong relationship between the skills taught and the literacy requirements of actual jobs."

It has been argued, especially by representatives of unions (Sarmiento & Kay, 1990), that workplace literacy programs that best serve both employer's and worker's interests should not focus on the requirements of specific jobs. Rather, broader competence should be sought so that workers can more productively switch from one job to another or from manufacturing in one way, using one set of tools, to another way that uses different tools and procedures.

Indeed, much of the training that companies are initiating today results from changes in the organizational approach to work. Under the "high performance," "focussed factory," or "TQM" models of work, workers must take on a broad range of tasks, including the ability to change from making one product today to making another within a very short turn

around. This creates the need for broadly conceived "workplace literacy" training that goes well beyond what a highly focussed "job-linked" program suggests.

In this case, whereas the training is "job-*linked*," in the sense that it relates to the person having a job position in a particular organizational setting, it is not "job *specific*" in the sense that it aims to teach only those reading, language, or mathematics skills identified by a "literacy audit" as those needed to perform a fixed set of tasks for a well-defined job position that includes a limited number of prescribed tasks.

It is possible to expand the concept of workplace literacy beyond the realm of a specific job or job position to encompass practically all education. This can be done through rationales such as, "If a worker is having a tough time parenting, then this could affect his or her work." Therefore, "Parenting Education" is a form of "job-linked" literacy. And, obviously, since elected officials of one philosophy are likely to be more supportive of work and workers than some others are, then "Civic Education" that explains our form of government, various ideologies, etc., is also "job-linked" literacy because how the workers vote may affect whether they will have a job.

If "workplace literacy" becomes identical to "adult education and lifelong learning," then the only remaining feature that distinguishes it from other programs is that it is education delivered at the workplace. And, indeed, organizations (including unions) that offer high school equivalency (General Educational Development - GED) programs at the workplace are operating from a loose definition of "job-linked" that argues that a high school diploma or its equivalency is a "ticket" to employment, and perhaps to increased productivity resulting from "general competence" or higher self-esteem, new found confidence, etc.

Because GED programs are not likely to be based on the literacy requirements of specific job tasks, the government guidance for the evaluation of the effects of workplace literacy programs on job productivity must be modified. It is possible that obtaining a GED will not improve job productivity - though if it does that is a finding that should be documented. Also, curriculum materials will likely not reflect the needs of the workplace. Rather, they will consist of GED "prep" materials. Assessments will consist of the GED tests (or other assessments, such as the External Degree in New York, leading to a high school diploma or its equivalence).

On what basis should companies decide whether to fund workplace literacy programs?

I base these comments on what I call the DOEED (pronounced, "do ed") approach to workplace literacy: *Developing Organizational Effectiveness* through *Employee Development*. From the DOEED perspective, companies should fund workplace literacy programs when they can determine that their organizational effectiveness will be improved by increasing their employee's development in literacy (more or less broadly construed).

In following the DOEED approach, a company analyzes how a commitment to employee development through literacy education might make functions such as: *Public relations* with the community more effective (here a company may decide to sponsor community literacy tutoring programs by local volunteer literacy groups, such as Literacy Volunteers of America, Laubach Literacy, etc., or it may contribute to pre-employment, job training programs to prepare youth and adults for work); *recruiting* more effectively from a broader pool of applicants by offering vestibule, job-linked literacy programs;*training* more effectively in the face of technological or organizational changes requiring upskilling; *performing* more effectively by increasing employee morale by providing education and

training opportunities; *promoting* deserving employees; and *outplacing* employees who may need workplace literacy training to find a new occupation.

The DOEED approach to workplace literacy broadens the approach of workplace literacy beyond a focus on improving the productivity of the presently employed. The DOEED approach looks at organizational effectiveness in a broad sense, not just in terms of productivity (though it does look at productivity, too).It considers the organization as a part of the larger community and as a contributor to the general welfare of the society. It considers that, in many respects, what is good for the community is good for the company.

For instance, it is considered to be in the long range best interest of a given company to help upgrade the skills of people being outplaced (laid off through "downsizing," "rightsizing," etc.). As an example, if an automobile manufacturing plant is being closed, it will be in the best interest of the company to provide workplace literacy training to prepare workers for other occupations. This can lead to higher employment rates, a larger tax base, better schools and a more highly skilled workforce pool to be recruiting from in the future. Chapter 3 provides an example of an evaluation study in the Chicago area that used the DOEED approach.

Developing an Attitude for Inquiry

As a final observation, it should be emphasized that one of the goals of evaluation is to permit the improvement of programs, not to simply decide if they work or not. The gathering of the types of information discussed in this chapter should be undertaken in the spirit of inquiry - always questioning, seeking information, and using that information to modify programs to make them more effective. Programs that seek to instill the love of lifelong learning in the workforce by starting learners off with the first steps into workplace literacy, should themselves exhibit positive attitudes toward learning - learning what they are doing, how they are doing it, and what might be done to improve what they are doing. Programs that hope to make critical thinkers of others should become models of critical thinking themselves. Good evaluation requires critical thinking, continuous learning, and thorough documentation to permit others to properly place a high value on good works.

References & Resources

Barker, K. (1990, December). *A Program Evaluation Handbook for Workplace Literacy.* Edmonton, Alberta, Canada: Kathryn Chang Consulting.(403) 437-4491.

Bonner, B. E. (1979, May). *A Survey of USAREUR Entry Level Skills of the 11B Infantryman.* Technical Report (TR-79-B3), Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Brown, E. (1990, June). *Evaluation of the R.O.A.D. to Success Program.* University Park, PA: The Pennsylvania State University, Institute for the Study of Adult Literacy.

Carnevale, A., Gainer, L. & Meltzer, A. (1990). *Workplace Basics Training Manual.* San Francisco, CA: Jossey-Bass.

Drew, R. & Mikulecky, L. (1988). *How to Gather and Develop Job Specific Literacy Materials for Basic Skills Instruction. A Practitioner's Guide.* Bloomington, IN: School of Education, Indiana University.

Faison, T., Vencill, M., McVey, J., Hollenbeck, K., & Anderson, W. (1992). *Ahead of the Curve: Basic Skills Programs in Four Exceptional Firms*. Washington, DC: The Southport Institute for Policy Analysis.

Kutner, M., Sherman, R., & Webb, L. (1990, October). *Review of the National Workplace Literacy Program.*. Washington, DC: U.S. Department of Education.

Mikulecky, L & Ehlinger, J. (1988). *Training for Job Literacy Demands: What Research Applies to Practice.* University Park, PA: The Pennsylvania State University, Institute for the Study of Adult Literacy.

Mikulecky, L. & D'Adamo-Weinstein, L. (1990, November). *Workplace Literacy Program Evaluations.* Scarsdale, NY: Work in America Institute.

Mikulecky, L. & Lloyd, P. (1993). *The Impact of Workplace Literacy Programs: A New Model for Evaluating the Impact of Workplace Literacy Programs.* Philadelphia, PA: National Center on Adult Literacy.

Nurss, J. R. (1990, March). *Hospital Job Skills Enhancement Program: A Workplace Literacy Project.* Atlanta, GA: Center for the Study of Adult Literacy, Georgia State University.

Philippi, J. (1991). *Literacy at Work: The Workbook for Program Developers.* New York: Simon & Schuster, Inc.

Sarmiento, A. & Kay, A. (1990). *Worker-Centered Learning: A Union Guide to Workplace Literacy.* Washington, DC: AFL-CIO Human Resources Development Institute.

Sticht, T.G. (1975, June). *A Program of Army Functional Job Reading Training: Development, Implementation, and Delivery Systems.* Final Report HumRRO-FR-WD(CA)-75-7. Alexandria, VA: Human Resources Research Organization.

Sticht, T. (1975). *Reading for Working: A Functional Literacy Anthology.* Alexandria, VA: Human Resources Research Organization.

Sticht, T. & Mikulecky, L. 1984). *Job-Related Basic Skills: Cases and Conclusions.* Columbus, OH: ERIC Clearinghouse on Adult, Career, and Vocational Education.

Sticht, T., Armstrong, W., Hickey, D., & Caylor, J. (1987). *Cast-off Youth: Policy and Practice from the Military Experience.* New York: Praeger.

Sticht, T. (1990, January). *Testing and Assessment in Adult Basic Education and English as a Second Language Programs.* Washington, DC: U.S. Department of Education.

Sticht, T. G. (1991, April). *Evaluating National Workplace Literacy Programs.* Washington, DC: U. S. Department of Education, Division of Adult Education and Literacy.

U.S. Departments of Education and Labor (1988).*The Bottom Line: Basic Skills in the Workplace.* Washington, DC: U. S. Department of Labor.

Chapter 3

Case Study Using the "DO ED" Approach for Evaluating Workplace Literacy Programs

Chapter 2 outlined the approach to workplace literacy program evaluation called Developing Organizational Effectiveness through Employee Development (DOEED) (pronounced "Do Ed"). This chapter illustrates how the DOEED approach was used to evaluate National Workplace Literacy Programs (NWLP) that were conducted in the Chicago area.

Education Partners

In 1992, the Workplace Education Division of THE CENTER/CCSD #54 of Des Plaines, Illinois, an educational agency, in partnership with the Management Association of Illinois (MAI) were awarded a National Workplace Literacy Program (NWLP) grant from the U. S. Department of Education. The grant was awarded to provide workplace literacy programs to industries in the Chicago area that were undergoing organizational changes to introduce one or more Total Quality Management (TQM) procedures.

Total Quality Management procedures typically involve the introduction of new skill demands on line employees. Though not all plants introduce all aspects of TQM, the procedures introduced generally result in changes in the ways that employees must work. Frequently employees must change from working alone to working in teams, they must change from performing limited functions to performing a number of different steps and operations to produce a completed product, they must change from having quality determined by an inspector at the end of a production line to building-in quality themselves by conducting various measurements and charting the results in what is known as "statistical process control-SPC," and they must frequently engage in more communications with customers. Additionally, in some cases the introduction of new technology requires that employees engage in training programs that are brief, intense and place a premium on good reading, studying, problem solving, mathematics and communication skills.

Business Partners

In the Chicago area, THE CENTER/MAI team became partners with ten businesses that were implementing one or more aspects of TQM. Through a preliminary needs assessment, it was determined that these industries had a combined workforce in which some 30% - 50% were lacking or weak in the basic English, literacy, or mathematics skills needed to work effectively in the new TQM environment. The businesses that were studied are briefly described below (note: these descriptions reflect the companies at the time of the origination of the project).

> **Amurol Product Company** manufactures speciality confectionery products. Of the 395 employees, there are 310 production workers on two shifts. In an effort to increase market share and due to the nature of business, new products are continually being introduced. Although the majority of sales are to domestic customers, new growth markets are being cultivated out of country.
>
> **Burgess-Norton Mfg. Co.** is involved in the development and manufacture of piston pins, shafts, powdered metal parts, casungs and keys, and sub-assemblies. These products are primarily produced for the automotive, truck and agricultural industries. A few of their major customers

include John Deere, Ford, General Motors, Caterpillar, and Chrysler. The company has been in business in Illinois since 1903 and currently employs 512 people at two locations.

Commander Packaging is a corrugated box manufacturer. The company has two plants in the Chicagoland area that employ 126 production employees who are members of the Graphic Communications Union. The company manufactures about a thousand custom orders each month. Their customers continue to demand more measurement and control of the manufacturing process. These demands result in more complex machinery, as well as a need for higher skill levels from all. The company is in the beginning stages of implementing Statistical Process Control in a plant-wide improvement process.

ITT McDonnell & Miller manufactures boiler feeders, water cutoffs, steam vents and pressure regulators. The company has a workforce of 300 employees with 170 in production; the majority of whom are members of the International Brotherhood of Boilermakers. In an effort to increase productivity, ITT has developed "production centers" and "focused factories." The next phase will be formalized SPC training for all employees.

John Crane, Inc. is a manufacturer of mechanical cells. Major customers include pump companies, the automotive industry, and other petroleum-related businesses. The company has a total workforce of 1,455 with approximately 841 involved in production. The company, in order to become more productive and increase its competitiveness, is employing the use of employee involvement and Statistical Process Control efforts, in order to increase employee effectiveness. In addition to the Total Quality Management, innovative work flow is being affected by the introduction of work cells.

Land O'Frost manufactures shelf stable food products and MRE (Meals Ready To Eat) for the military and was one of the primary food providers for Operation Desert Storm. The company has a total workforce of 275 which includes 225 production employees who are members of the United Food Commercial Workers.

Parco Foods, Inc. is a leading baker of specialty cookies in the United States. The company supplies baked and frozen dough to a wide variety of wholesale and institutional distributors, as well as retailers of cookies such as MacDonald's. Approximately 211 members of the General Service Employees Union are employed on a full-time basis with up to 100 additional individuals employed seasonally.

Phoenix Closures, Inc. develops, manufactures and markets closures, fitments and container sealing systems used in packaging a wide range of consumer, industrial and institutional products. Since 1982 the company has manufactured thermoplastic caps exclusively. The employment at Phoenix Closures has stabilized as their market matured so that nearly 300 individuals are employed today. Of that total, 208 are members of the Amalgamated Clothing & Textile Workers Union. In an effort to remain competitive, the company modernized processes and developed new products, as well as initiated a Total Quality Management program.

Tricon Industries, Inc. is manufacturer of custom inserted molded components for the automotive industry and switches for the appliance industry. Since the company was started in 1944, it has expanded to 340 employees in four locations. Over the past two years Tricon has experienced significant growth in direct labor positions and support personnel.

Videojet Systems International is a subsidiary of A. B. Dick Company. The company manufactures continuous stream ink jet processing printers and specialty inks. The production force totals about 270. The company has plans to implement SPC and an overall employee involvement initiative.

Meeting the Needs for Workplace Literacy

The preliminary analyses of the needs for basic skills training in the ten Chicago-area industries revealed that the primary needs were those for English language training, reading and writing literacy skills, and numeracy (computation, graphs) skills.

Establishing Workplace Literacy Programs

To establish basic skills programs, each industry training site established its own Employee/ Employer Basic Skills Committee. Each committee was comprised of a Human Resource Development/Personnel staff member, a plant manager, a floor supervisor, the union President or shop steward (if unionized), at least two production employees participating in the program, and a Site Coordinator.

The Committee made joint decisions on each aspect of the program design and implementation, including:

* a recruitment plan
* assessment policy and selection of assessment instruments
*review of overall assessment statistics
* approval of the course schedule and curriculum
* evaluating the achievement of program outcomes
* participation in the evaluation of the impact of the Basic Skills Program

Job Basic Skills Course Curriculum Development. To meet the specific basic skills needs of each of the ten industries, THE CENTER/MAI team produced customized training programs that were based on discussions with supervisors and employees regarding the specific types of job tasks that were producing some difficulties for workers because of basic skills problems. Additionally, an analysis was made of the types of tasks related to TQM that employees at each company had to perform that involved the use of English, reading and writing, and/or mathematics.

Observations of employees at work were accomplished to determine how basic skills were used on the job. Copies of job materials, including materials used in job training programs were obtained and were used to develop job-related curriculum materials. These materials included lists of the competencies that were to be developed, job-related basic skills tests that could be used as pre- and post-tests to determine if what was taught was learned by employees, and course materials used in instruction and for learning by employees.

Accomplishments

Number of Courses Conducted. Though the THE CENTER/MAI programs were originally supposed to extend for only six quarters, an extension was obtained from the U. S. Department of Education that permitted two extra quarters in which courses could be presented.

Altogether, a total of 104 courses was offered in the project, which is about 108% of the total of 96 courses that was originally estimated to be needed. Most of the courses ran for 36-40 hours. They were offered on company time for the most part , though in some cases employee time before or after work, or during lunch was used for half the course. Classes were held in meeting rooms provided by the company. The number of courses offered by each company was (from highest to least number of courses): Tricon Industries (22 courses); John Crane (18 courses); Burgess-Norton-16; ITT McDonnell & Miller-13; Phoenix Closures-9; Amurol Products-8; Land O'Frost-8; Videojet-7; Parco Foods-2;

Commander Packaging-1. Thirty-three of the courses were for English as a Second Language (ESL), 28 were for reading/writing, 35 were for mathematics, 6 were for preparation for the high school equivalency examination (the GED), and 2 were communications courses called "Customer Interaction."

Number and Costs of Employees Receiving Instruction. The data in this section is taken from the final quarterly report for the project. It shows that a total of 3,291 employees were assessed for basic skills across the ten industries and across all eight quarters of the project. This is 127% of the proposed goal of 2600 to be assessed. However, while the assessments exceeded the projected numbers, the courses actually enrolled only 948 employees, about 62% of the 1,525 that had been established as the goal for the project when originally proposed to the U. S. Department of Education.

Of the 948 employyes who participated in courses, their average age was 41 years, 45% were males while 55% were females. For those reporting race/ethnicity data, 29% were White, 8% were Black, 49% were Hispanic and 14% were Asian/Pacific Islander.

The cost of the project in federal funds was $455,607. For the 948 employees, this comes to $480.60 per employee student. When the additional in-kind funds ($120,839) are added to the federal costs, the sum is $576,446, or $608.07 per employee. Finally, when the value of the release time that companies provided is added to the previous costs, the total is $814,541 or $859.22 per employee.

A total of 21,289instructional hours were provided at a cost of $21 per hour in federal funds, and $38 per hour when all funds are considered. On the average, since each worker received about 22.46hours of instruction (21,289/948=22.46), the federal costs per employee were $471.66 and total costs were about $853.48 per worker.

Evaluating the Workplace Literacy Programs

Evaluation of the THE CENTER/MAI workplace literacy programs was accomplished by both internal and external evaluation activities. In the internal activities, the Project Director at THE CENTER was responsible for obtaining and reporting all of the data presented above on numbers, types, and costs of courses. The Project Director was also responsible for supervising the quality of all aspects of the various program start-up, development, implementation and reporting activities. The Project Director, working with staff, was also responsible for obtaining all the pre- and post-test data and for administering and recording the interview questionaires used to determine employer and employee perceptions of the workplace literacy courses.

The external evaluation activities consisted of site visits by the external evaluator to some of the locations and classrooms where instruction was carried out. This permitted the external evaluator to verify, on an unsystematic sampling basis, that quality instruction was being offered and that employers and employees were able to make judgments regarding the benefits of the instruction to them and the company.

In evaluating the workplace literacy programs, there were two main bodies of information that were developed. One dealt with how the program contributed to the organizational effectiveness (OE) of the business or industry involved in the program, and the other involved the effects of the program on employee development (ED).

The OE Perspective

From the perspective of the employing organization, workplace literacy programs are implemented to improve the organization's performance of one or more of its major human resources functions. These functions include public relations, recruitment, training, employee behavior, productivity (job performance) monitoring and improvement, and advancement and promotion of effective employees.

Table 3.1. Responses of supervisors to interviews regarding the effects of the workplace literacy programs on organizational effectiveness in various human resources functions.

Organizational Effects

Company	Public Relations			Recruit Employees			Easier Training			Improved Employee Behavior		
	Yes	No	DK	Yes	No	DK	Yes	No	DK	Yes	No	DK
Amural		1	2	1		2	2		1	3		
Burgess-Norton		1	1		2		2			1	1	
John Crane		1	2	1	1	1	2	1		2	1	
ITT M & M		1	1	1		1	1	1		2		
Phoenix Closures			4		1	3	3	1		2	1	1
Tricon			3			3	2		1	3		
Videojet			4			4	1		3	2	2	
Totals	0	4	17	3	4	14	13	3	5	15	5	1

Organizational Effects

Company	Productivity			Promotions			Other Effects			Continue Program?		
	Yes	No	DK	Yes	No	DK	Yes	No	DK	Yes	No	DK
Amural			3	1			2	1		2	3	
Burgess-Norton				2			2	2			2	
John Crane	2		1	1		2	2	1		1	2	
ITT M&M	1	1	2		2		2					
Phoenix Closures	2	2		3	1		4			3	1	
Tricon	2		1		2	1	3			1	2	
Videojet	3		1		4	3	2				4	
Total	10	3	8	7	9	5	17	0	4	7	0	14

In evaluating the workplace literacy programs, the external evaluator designed interviews that were administered to an unsystematic, convenience sample (obtained by the Project Director) of managers and supervisors to determine whether in their judgment, the workplace literacy programs had contributed to one or more of these organizational functions.

Table 3.1 summarizes the Organizational Effectiveness interviews for seven companies for which a total of 21 interviews were conducted by the THE CENTER staff. The remaining four companies were not sampled due to the time and expense involved in making numerous appointments and then re-scheduling when supervisors and/or employees could not make previously scheduled meetings. Repeated cancellations of scheduled meetings occured because of business factors even when the external evaluator had traveled to the Chicago area with previous appointments made.

*Public Relations and Recruitment Functions.*The combined data indicate that, for the most part, the supervisors interviewed were unaware of whether or not the programs had helped the companies' public relations (e.g, through newspaper stories or company newsletters) or employee recruitment functions. Three supervisors,at Amurol, John Crane and ITT M&M thought that the programs had improved their companies' ability to recruit new employees. The supervisor at John Crane thought this was so because the company offered workplace literacy programs now. Presumably, this would permit John Crane to recruit from a larger pool because it would not have to reject as large a number of less literate applicants.

Training Function. Two-thirds of the supervisors thought that the workplace literacy programs had improved their companies' ability to conduct training. Specific comments included:

> Burgess-Norton: (1) "Math classes will help with SPC; English classes will help with team training; employees more confident." (2) "Should help with SPC training."
>
> John Crane: (1) "They're capable of training their co-workers." (2) "Better communication."
>
> ITT M&M: (1) "Basic skills will help them with training."
>
> Phoenix Closures: (1) "Easier to train."(2) Some employees easier to train." (3) "Easier to train."
>
> Tricon: (1) "Easier than before - pay more attention to details."
>
> Videojet: (1) "Helped with other classes."

Employee Behavior. Seventy-one percent of supervisors thought that the workplace literacy programs had affected employee behaviors on the job. Specific comments included:

> Amurol: (1) "People participating in program were more involved becasue they could communicate more ideas." (2) Employees have displayed some improved satisfaction that company has made an effort to provide help." (3) "Participants have exhibited an increase in self image which in turn has helped them in teamwork, helping in a positive manner in all work related duties."
>
> Burgess-Norton: (1) "Speak more."
>
> John Crane: (1) "---- has improved a bit. She's more confident now than before. ---- is about the same." (2) "Morale & teamwork is rising due to the increased confidence in communications."
>
> ITT M&M: (1) "Improved attitude about the company-people seeing company doing something for them." (1) "A greater willingness to write out ideas, less afraid."

Phoenix Closures: (1) Teamwork improved."

Tricon: (1) "Increase morale, confidence to participate in teams." (2) "Morale." (3) "Morale higher."

Videojet: (1) "Some improvement." (2) "Understands better."

Productivity Function. In some cases the workplace literacy program may help improve an employee's job productivity through the reduction of errors, wastage, or other such efficiencies. In the present case, over one-third (36%) of the supervisors interviewed stated that they thought the workplace literacy programs had helped improve productivity in one way or another.

John Crane: (1) "Rising levels of effective communication is reducing the amount of scrap."

ITT M&M: (1) "More accuracy in reporting."

Phoenix Closures: (1) "Some, not all employees improved productivity." (2) "Less scrap."

Tricon: (1) "Reduce errors paperwork." (2) "Better on paperwork. Fewer errors paperwork. More conscientious."

Videojet: (1) "Understands and asks questions more now."

Promotion Function. At times, employee's basic skills levels may be too low for them or the company to consider them for promotion. In the present project, five supervisors in three companies thought that for some employees, their participation in the workplace literacy programs had increased their chances for promotion.

Amurol: (1) "This is too early to evaluate at this time. ---- was a back up line leader and more fully utilized as a line leader. The improved skills were of some assistance."

John Crane: (1) "In case there will be an opening, ----is qualified to be promoted."

ITT M&M: (1) "Trap line is more self-reliant, less dependent on salaried people." (1) " It hasn't happended yet because there isn't much movement, but he predicted people will be easier to train."

Phoenix Closures: (1) "Potential to promote."(2) "One may be ready to promote."(3) Some have promoted. Some will."

Other Effects. In almost four out of five cases (80%) the 21 supervisors who responded to the organizational effectiveness interview stated that there were other effects that the workplace literacy programs had had in addition to those previously discussed. Specific comments included:

Amurol: (1) "Safety-helped people to read important signs & machinery parts; Data Collection-helped people understand appropriate paperwork; Communication-with supervisors improved."

Burgess-Norton: (1) "One communicates more now wtih supervisors. Supervisors more confident employees understand instructions." (2) "Positive attitude-liked class or getting off work."

John Crane: (1) "I've noticed that most workers who participated improved their self confidence, speaking and working." (2) "Employee confidence-better command of speaking/writing; Employee participation increased-result of confidence; Empowerment & team building can be focused on.

ITT M&M: (1) "Positive attitude-people appreciate it & feel better about the company." (1) Classes have helped people understand information at work & indirectly ISO 9000."

Phoenix Closures: (1) "Spelling improvement; Involvement in meetings increased." (2) "Enthusastic about learning." (3) "More willing talk at meetings." (4) "More aggressive about jobs-try improve their skills."

Tricon: (1) "In promotable status-some participants will be more likely to promote than before."(2) "Self-esteem improved." (3) "Better understanding-speak better (ESL students); Math better for SPC."

Videojet: (1) "Eager - talk to others - 1 especially." (2) "Took shyness away." (3) "Not afraid to communicate now; Takes more initiative-starts on own."

Will the Company Continue the Program? This question was included to get yet another indication of the extent to which companies valued the workplace literacy programs. It is not likely that companies would want to continue programs that they did not feel were valuable.

In the present case, seven (33%) supervisors at four companies stated that they thought the company wanted to continue the programs. Specific comments included:

Burgess-Norton: (1) "Planning to continue beyond grant. Prefer 1/2 on company time, 1/2 on employee time because of impact on production schedule." (2) "Committed to continuing on own. Took longer for empoyees to reach goals than he anticipated. Apprehension about the classes has subsided."

John Crane: (1) " We are looking into a state grant."

Phoenix Closures: (1) "Would like to see training continue. Will be more training (union will be conducting training)." (2) & (3) "Will continue (union will be conducting training). Think good idea to continue.

Tricon: (1) "Math training-positive & negative numbers."

Summary of the OE Responses. Summing across the "Yes," "No," and "Don't Know" columns of Table 1 gives 72 "Yes," 28 "No,' and 68 "DK" responses. If attention is restricted to only the "Yes" and "No" responses, there were a total of 100 responses, of which 72% were "Yes," indicating that the program has had a positive effect on one or more organizational human resources functions.

While the interviews were open-ended and permitted supervisors much leeway in responding, the fact that so many "Don't Know" responses were recorded suggests that supervisors were not responding to the interview with a simple bias toward positive responses. Rather, they seemed to be reluctant to comment when they felt that they did not know enough to comment.

That so many of the supervisor's responses commented on the new found confidence and self-image of employees is a perception that they shared with the employees themselves, as was indicated in the employee development interviews summarized later on.

The ED Perspective

While the OE perspective places the needs of the organization at the forefront of program evaluation, the employee development (ED) perspective looks at how the program is serving the interests of the employee in both the workplace and in other settings.

Becoming involved in a job- based education program can motivate employees to seek more responsibility at work, it can affect their attitudes toward schooling and learning, and this can affect their behaviors toward their children, spouses and others. It can motivate employees to continue their education outside of the workplace. All these changes can, in turn, increase the "marketability" of the person and influence supervisors and managers to a greater appreciation of the person as an employee, and this may be reflected in increased pay and promotions or a job change. These types of employee developments serve to indicate that the workplace literacy program has produced a degree of "portability" of literacy skills in the employee.

Learning Outcomes

The first type of information that is useful in determining ED effects is information about how well the employees learned in the various courses. Information regarding learning outcomes were obtained by the internal evaluation staff. This information included data on the percentages of enrollments, drop outs, and success rates of those who completed the various courses. Additional information was obtained using job-related English, reading/writing or math tests that were administered as both pre- and post-tests to measure the extent to which employees learned what was taught in the courses. Pre- and post-test data from courses in six companies were provided to the external evaluator for analysis and reporting.

Course Completion and Success Rates. Of the 948 employees who participated in the 104 workplace courses, 33% were enrolled in ESL programs, 34% in Math, 26% in Reading/Writing, 5% in GED preparation, and 2% in Customer Interaction programs. There was an 11% drop out rate across all programs. For the 89% who remained in the programs, there was a 95% success rate in which employees met the standards for mastering the competencies taught in the courses. The standards for the competency-based courses was that at the end of the course, 90% of employees will demonstrate the competencies taught in the course.

Demographics of Employees With Test Score Data. To determine if employees had learned what was taught in the job-related reading and mathematics courses, tests were constructed using job materials and asking for task performance similar to that needed for reading or computing on the job. Only one form of each test was constructed. It was used for both pre- and post-testing. It was expected that because there were several weeks and some 38 or so hours of instruction between the pre- and post-tests that the gains exhibited would reflect learning due to instruction and not just practice in taking the test once before taking it again. The procedure of constructing alternate forms of tests for pre- and post-testing that were psychometrically equivalent was too technical for the internal evaluator staff and would have been too costly for the project's budget if tests had been developed by either internal staff or external consultants. It would also have demanded considerable participation by employers and employees beyond that which was devoted to instruction, and such additional time and personnel commitments from the industries involved were not feasible.

In the case of the mathematics tests, they were decontextualized problems in computational operations (add, subtract, multiply, divide) taken from the Tests of Adult Basic Skills (TABE). Because the tests were excerpts and not complete tests, use of the norming data for the TABE was not appropriate.

Table 3.2 shows data from eleven courses conducted at six companies. Demographic data for each company is summarized in the following.

Burgess-Norton: Data for one reading and one mathematics course were available from Burgess-Norton. There were 9 employees in the reading course, all of whom were ESL students. Eight were male and all were Hispanic. Ages ranged from 29 years to 57 years, with a mean of some 40 years. Four had 6 years of education, one 10 years and 3 had completed 12 years of education. They had been employed from 1 to 17 years, with 1 year being the median.

For the 35 members of the mathematics course, 28 (80%) were males, and 9 were ESL. Thirteen (37%) were White, 13 (37%) were Black, and 9 (25.7%) Hispanic. Their ages ranged from 25 to 60 years, with an averge age of 40 years. Ten were 45 years old or older. Their years of education ranged from 8 to 12, with over 18 having 12 years of education. They had been employed anywhere from 1 to over 21 years, with 26 (74%) having been employed 10 or fewer years. Only two had been employed for less than one year. The median years of employment was 6.

John Crane: Data from one reading and one mathematics course were available from John Crane. Of the 16 employees in the reading course, 9 were male and all were ESL language users. There were no Whites or Blacks in the program. Regarding ethnicity, there were 6 (37.5%) Hispanics, 4 (25%) Asian, and 6 (37.5%) Other. Their ages ranged from 30 to 64, with an average of 46 years. Nine were 45 years old or older. Their years of education ranged from 5 to 13, with 6 having 12-13 years of education. The median years of education was 9.5. They had been employed from 6 to 20 years, with a median of 12.5 years of employment.

Of the 14 employees in the mathematics program, 13 were female, and 13 were ESL speakers. There were no Whites, there was 1 (7%) Black, 2 (14.2%) Hispanics, 1 (7%) Asian, and the remaining 10 (71%) were Other. Their ages ranged from 25 to 67 years, with an average of 42 years. Their years of education ranged from 4 to 8 years, with a median of 4.5. They had been employed from 4 to 23 years, with the median years of employment being 6.

ITT McDonnell & Miller: There were 21 employees in the reading program for which data were available. Fifteen of the employees were males, and 19 were native English speakers. There were 10 (47.6%) Whites, 7 (33.3%) Blacks, and 4 (19%) Hispanics in the class. Ages ranged from 34 to 63, and the median was 48 years of age. Nine were over 50 years of age. Their years of education ranged from 3 to 17, with 11 having 12 or more years of education. The median was 12 years of education. They had been employed from 1 to 27 years, with a median of 13 years of employment.

Phoenix Closures: Data were available for one reading and one mathematics course at Phoenix Closures. In the reading course, there were 13 employees, of whom 7 were females, and 10 were ESL speakers. Four (30.7%) were non-Hispanic Whites, and the remaining 9 (69.2%) were Hispanic. Ages ranged from 24 to 45 years , with a mean age of 37 years. Years of education ranged from 6 to 12, with a median of 9 years. Years of employment ranged from just over a half year, to 13 years, with a median of 6 years.

In the mathematics course, there were 38 employees who participated. Six of these had also taken the reading course. Of the 38 employees in the course, 13 were males and 25 females. Sixteen were native English speakers and 22 were ESL speakers. Eighteen (47%) were non-Hispanic Whites, 19 (50%) were Hispanics, and 1 was Asian. Age data were available only for the six employees who had taken the reading course, and ages ranged from 28 to 44 with 4 being over 40 years of age. Years of education ranged from 4 to 12, with a median of 9. Years of employment ranged from 2 to 11, with a median of 5 years.

Tricon: Data were available for three courses at Tricon, two reading and one mathematics course. One reading course was for employees in general, and the second was only for employees in the production division of Tricon. Demographic data were available only in the course for general employees. In this course, 17 of the 19 employees were female and were ESL speakers. Ages ranged from 26 to 52, with a median of 35 years of age. Two (10%) were White, 3 (15.7%) were Black, 8 (42%) were Hispanic, and 5 (26%) were Asian. Years of

education ranged from 6 to 16, with a median of 11. Median years of employment was 1.5, with a range from 0.2 to 14 years.

In the mathematics class there were 11 employees, 9 or whom were female. Four were ESL speakers. of whom 7 (63.6%) were White, 2(18%) were Black, and 3 (27.7%) were Asian. Years of education ranged from 8 to 12, with 8 having 12 years of education. The median age was 47, with the range going from 34 to 54 years. Years of employment ranged from 0.8 to 14, with the median being 3 years.

Videojet: The 15 employees in the reading program with data from Videojet were 7 males and 8 females, all of whom were native language speakers. Six (40%) were Black, 6 (40%) were Hispanic, and 2 (13.3%) were Asian. Years of education ranged from 8 to 16 years, with a median of 12, and years of employment ranged from 2 to 16, with a median of 6.

Pre- and Post-Test Scores. It is clear from the mean scores of Table 3.2 that in all cases, employees did considerably better on the post-tests than they did on the pre-tests, suggesting that all courses resulted in learning by the participants. Indeed, out of the total of 209 pre- and post-test scores across all courses and companies, 207 showed positive gains and only two showed post-test scores lower than pre-test scores, and both of those were in the mathematics tests which were multiple-choice and permitted guessing.

Table 3.2. Means and standard deviations (SD) of pre- and post-test scores on job-related reading and math tests in eleven courses at six companies. All entries are raw scores correct except for John Crane-Reading which are percent correct. All pre-post gain differences are statistically significant using t-tests for paired means.

Company	N	Reading			N	Math				
		Pre		Post			Pre		Post	
		X	SD	X	SD		X	SD	X	SD
Burgess-Norton	9	18.7	19.7	32.0	12.3	35	21.7	28.3	28.3	11.3
Max. Possible:		47				44				
John Crane	16	44.3	22.2	70.8	18.9	13	26.3	04.8	34.3	06.7
Max. Possible:		100%					48			
ITT	21	29.2	10.7	40.9	06.1	-	-	-	-	-
Max. Possible:		56								
Phoenix	13	45.7	16.5	99.5	08.4	38	28.9	06.6	38.4	05.6
Max. Possible:		125					48			
Tricon	19	35.3	09.3	58.9	10.4	11	18.0	07.5	27.9	05.8
Max. Possible:		74					34			
	19	11.1	04.6	18.1	02.9	-	-	-	-	-
Max. Possible:		21								
Videojet	15	38.5	05.4	52.3	03.1	-	-	-	-	-
Max. Possible:		62								

Employee Interview Responses. The test score data indicated that employees did, in fact, learn job-related knowledge in the courses they attended. However, some literacy educators have speculated that workplace literacy programs that focus on job-related knowledge may result in learning that has little or no transfer, "portability,"or generalizability to situations outside the workplace.

To get some idea about how employees felt about the value of the workplace literacy programs for work, home and community, the structured interviews asked for detailed information as indicated in Tables 3.4, 3.5, 3.6, 3.7 and 3.8.

Table 3.3 presents a summary of the responses from the 22 employees interviewed in four companies. Clearly, the workplace literacy programs were not viewed as entirely restricted to helping the employees at work. Summed over the four companies, more than half thought that the programs not only helped them at work, but also at home. Some 40% thought the programs had helped them in their communities.

Table 3.3. Employee responses to interviews about how the workplace literacy programs had helped them.

		Has This Workplace Literacy Program Helped You At:											
Company	N	Work			Home			Community			More Education		
		Yes	No	DK	Yes	No	DK	Yes	No	DK	Yes	No	DK
		%	%	%	%	%	%	%	%	%	%	%	%
Burgess-Norton	5	76	16	8	64	18	18	33	67	0	80	0	20
John Crane	8	65	27	8	33	67	0	69	31	0	50	37	13
ITT M&M	5	91	7	2	64	36	0	95	5	0	100	0	0
Tricon	4	76	21	3	58	42	0	25	75	0	75	25	0
Videojet	5	62	19	19	58	42	0	55	35	10	40	60	0

Note: This table shows the percentage of Yes, No, or Don't Know responses to questions about the effects of participating in workplace literacy programs on work, home, community, or desire for additional education. For instance, considering John Crane, there were 8 employees who answered 10 questions about the effects of the program on work. Thus there might have been 80 responses. However, because one of the questions was about a math program, and none of the employees at John Crane took a math program, the math question was not applicable to these eight students. Therefore the potential of 80 responses was reduced by 8 to 72. Then, because a second question on teamwork was not applicable to these 8 employees, because they all worked alone, the potential of 72 responses was reduced by 8 to 64. The table shows the percentage of the 64 remaining responses that were Yes, No, or DK responses. For John Crane, 65% of the 64 responses were Yes, 27% were No, and 8% were DK. Similar procedures were followed in constructing the remaining data in the table.

Contributions to National Education Goals. National education goal number 6 (in the Goals 2000 Act) calls for adults to engage in lifelong learning. Importantly, over half of the employees stated that their participation in the workplace literacy program had stimulated an interest in participating in additional education, suggesting that the programs have contributed to the achievement of goal number 6.

National education goals number 1 states that all children will enter school ready to learn, and it places quite a bit of responsibility upon parents or grandparents for preparing their children for school by reading to them during the pre-school years. Examination of Tables 3.4, 3.5,3.6, 3.7 and 3.8 reveals 12 of the 22 respondents had no children or grandchildren to read to. But of the remain 10 employees, 40% said that due to the workplace literacy program they now read more to their children. This suggests that the workplace literacy programs may also contribute to the achievement of Goal 1.

Table 3.4. Employee Development Effects

Burgess-Norton

Has this ESL/Read &Write or Math program helped you at work:	Yes	No	DK	Example/Comment
1. Read job materials better?	3			
2. Write job materials better?	3			Some words; Sometimes-more words
3. Listen & speak on the job better?	3			Understand more now. Understands verb tenses. Speaks more; understands better now.
4. Do math for job tasks better?	2			Refresh memory; Better understanding now.
5. Work better in teams?				(n/a - all work by themselves)
6. Reduce waste; scrape; errors; etc. ? 3	2			(1) Less errors in paperwork.
7. Know more about company policies, etc.	2	1		(2) Understands better now.
8. Feel confident about trying for promotion?		1	2	(1) Maybe later.
9. Learn better in company training programs?				(n/a- none have taken other training)
10. Improve your morale with company?		3		
Has program helped you at home?				
11. Have you started reading more at home?	3			Uses dictionary to read paper in English. Does homework for community college course. Reads little more now.
12. Do you write more/better at home?	1	2		Before couldn't write anything.
13. Do you use math better at home?	2			(1) More comfortable now.
14. Do you help your children/ grandchildren with homework more?	1			Has daughter in 4th grade-help each other. (4 n/a)
15. Do you read to (grand)children more?		2	(1 n/a)	
Has this program helped you in your community?				
16. Do you feel more confident about reading in stores, offices, etc.		3		No problems with this type of reading.
17. Do you feel more confident writing in government forms, etc.		3		Usually have forms in Spanish too. No problems in this area.
18. Has this program made it easier for you to speak in public?	3			More confident; tries more.Depends on conversation. More comfortable now.
19. Has this program made you feel more confident about reading and understanding the issues for voting in the next election?	1	2		Not citizen; thinking about becoming citizen. Not too much-use different words.
20. Has the program lead you to consider taking more education or training programs?	4		1	Studies with videos at home-Spanish/English. Taking community college class-ESL Maybe weekends. Baby sits during week. Time problems.

Table 3.5. Employee Development Effects

John Crane

	Yes	No	DK	Example/Comment
Has this ESL/Read &Write program helped you at work:				
1. Read job materials better?	7	1		Read job forms better
2. Write job materials better?	3	4	1	Short sentences
3. Listen & speak on the job better?	8			Not ashamed now. Speak better.
4. Do math for job tasks better?				(n/a)
5. Work better in teams?	7	1		Easier to understand others.
6. Reduce waste; scrape; errors; etc. ?	1	5		A little.
7. Know more about company policies, etc.	8			(8) Understands safety better now.
8. Feel confident about trying for promotion?		6	2	Need more English. Too old. Need to read better.
9. Learn better in company training programs?				(n/a- none have taken other training)
10. Improve your morale with company?	8			Can talk to boss better. Very Happy.
Has program helped you at home?				
11. Have you started reading more at home?	4	4		Read paper at lunch time. (2) Read paper.
12. Do you write more/better at home?	1	7		Notes to daughter.
13. Do you use math better at home?				(n/a)
14. Do you help your children/ grandchildren with homework more?	1	3		(4 n/a) Help more with math than with English.
15. Do you read to (grand)children more?	2	2		(4 n/a)Reads to child. Reads when babysitting.
Has this program helped you in your community?				
16. Do you feel more confident about reading in stores, offices, etc.	8			Read signs better.
17. Do you feel more confident writing in government forms, etc.	5	3		(2) Driver's license. Fill forms out better.
18. Has this program made it easier for you to speak in public?	6	2		
19. Has this program made you feel more confident about reading and understanding the issues for voting in the next election?	3	5		Not citizen.
20. Has the program lead you to consider taking more education or training programs?	4	3	1	Like to try. If didn't have child. Computer classes.

Table 3.6. Employee Development Effects

ITT McDonnell & Miller

Has this ESL/Read &Write/GED program helped you at work:	Yes	No	DK	Example/Comment
1. Read job materials better?	5			Understand gauges & work orders better. Easier to read words & expresses self better.
2. Write job materials better?	5			Can fill out work order & tickets better. Helped fill out papers better w/fewer errors.
3. Listen & speak on the job better?	4			(1 n/a) Tremendous difference. Less shy; voice better. Can use more words. More ability to explain how work should done. Understand English better.
4. Do math for job tasks better? (n/a)				
5. Work better in teams?	5			Listens to others more. Hear their opinion. More considerate now of others. More able to understand other people & express his thoughts. Can explain better. Communicates better with different people.
6. Reduce waste; scrape; errors; etc. ?	4	1		Wastes less time now when writing. More thorough now & has a better work ethic. Helped him become neater. Can read instructions better which helps reduce scrap.
7. Know more about company policies, etc.	5			(5) Read & understand rules/policies better now.
8. Feel confident about trying for promotion?	5			Became a group leader! Confident he knolws his job well & can do any job. Made him more confident of reading ability "to handle different situations." Feels he is able to achieve in a harder job.
9. Learn better in company training programs?	4		1	He slows down and reads more carefully. Reads directions better. Can listen better & apy better attention. Gets along better w/people from different cultures. Works better w/people; better communication.
10. Improve your morale with company?	3	2		Feels better at work. Felt good that company offered him a program. Felt encouraged to write.
Has program helped you at home?				
11. Have you started reading more at home?	3	2		Helps wife with schoolwork. Newspapers & Bible.
12. Do you write more/better at home?	4	1		Better penmanship & spelling. Writes down fishing conditions for future reference. Starting to write checks & pay bills more. Writes notes from Bible to show his father. Writes about what other countries are producing on their farms.
13. Do you use math better at home? (n/a)				
14. Do you help your children/ grandchildren with homework more?	1	1		(3 n/a) Helps daughter with reading.
15. Do you read to (grand)children more?	1	1		(3 n/a)
Has this program helped you in your community?				
16. Do you feel more confident about reading in stores, offices, etc.	5			Read labels easier. Understand medical forms better.
17. Do you feel more confident writing in government forms, etc.	5			Able to explain himself better. Filled out a car registration last night. (2) Fill out forms better.
18. Has this program made it easier for you to speak in public?	4	1		More comfortable/confident. Thinks before speaks. Less shy.
19. Has this program made you feel more confident about reading and understanding the issues for voting in the next election?	4			(1 n/a - not citizen). Read/listen to news better.
20. Has the program lead you to consider taking more education training programs?	5			More job-related schooling. Taking courses for stationary engineers license. Pursue writing. Improve English with private tutor. Community college GED possibly.

Table 3.7. Employee Development Effects

Tricon

Has this ESL/Read &Write program helped you at work:	Yes	No	DK	Example/Comment
1. Read job materials better?	4			Terminology clearer. (2) Read forms better.
2. Write job materials better?	3	1		Lots better.
3. Listen & speak on the job better?	2	1	1	More sure of what said. Very improved.
4. Do math for job tasks better?				(n/a)
5. Work better in teams?	4			(3) Communicate better with others.
6. Reduce waste; scrape; errors; etc. ?	3	1		(3) Less mistakes with paperwork.
7. Know more about company policies, etc.	4			(4) Understand policies better now.
8. Feel confident about trying for promotion?	1	3		Would like to apply for better job.
9. Learn better in company training programs?	1			Took SPC class. Understood paperwork better. (3 n/a- have taken no other training)
10. Improve your morale with company?	3	1		(2) Feel better about self.
Has program helped you at home?				
11. Have you started reading more at home?	1	3		Reads bills better.
12. Do you write more/better at home?	2	2		Try write more. Writes notes to teacher.
13. Do you use math better at home?				(n/a)
14. Do you help your children/ grandchildren with homework more?	2			(2 n/a) Daughter helps her too.
15. Do you read to (grand)children more?	2			(2 n/a) Reads to her little boy. Easy to read children's books.
Has this program helped you in your community?				
16. Do you feel more confident about reading in stores, offices, etc.	2	2		(2) Goes self now, before needed interpreter/help.
17. Do you feel more confident writing in government forms, etc.		4		
18. Has this program made it easier for you to speak in public?	2	2		Lot more comfortable now. More confident.
19. Has this program made you feel more confident about reading and understanding the issues for voting in the next election?		4		
20. Has the program lead you to consider taking more education or training programs?	3	1		Might take classes at community college for better /different job. Maybe to learn more English.

Table 3.8. Employee Development Effects

Videojet

Has this ESL/Read &Write program helped you at work:	Yes	No	DK	Example/Comment
1. Read job materials better?	4	1		
2. Write job materials better?	3	2		Understands paperwork more. Fills forms more.
3. Listen & speak on the job better?	5			Understand better.
4. Do math for job tasks better?				(n/a)
5. Work better in teams?	4		1	(2) Understand more now. Little better now.
6. Reduce waste; scrape; errors; etc. ?	1	2	2	Less mistakes with paperwork.
7. Know more about company policies, etc.	4	1		(2) Understand rules/policies better now.
8. Feel confident about trying for promotion?	2		2	
9. Learn better in company training programs?	2			Two took other classes but ESL class didn't help. (3 n/a- have taken no other training)
10. Improve your morale with company?	2		3	A little. More comfortable speaking now.
Has program helped you at home?				
11. Have you started reading more at home?	3	2		(2) Newspapers. (1) magazines. Understands more.
12. Do you write more/better at home?	2	3		Write notes to kids, husband. Writes short notes.
13. Do you use math better at home?				(n/a)
14. Do you help your children/ grandchildren with homework more?	2			(3 n/a) Helps 8 year old. Help each other.
15. Do you read to (grand)children more?				(5 n/a- no little children/grandchildren)
Has this program helped you in your community?				
16. Do you feel more confident about reading in stores, offices, etc.	4	1		Don't always understand, but asks questions.
17. Do you feel more confident writing in government forms, etc.	1	4		
18. Has this program made it easier for you to speak in public?	4	1		More comfortable. Ask questions.
19. Has this program made you feel more confident about reading and understanding the issues for voting in the next election?	2	1	2	Likes to read about politics.
20. Has the program lead you to consider taking more education or training programs?	2	3		Like to take more classes at school. Has taken more classes outside work.

Conclusions and Recommendations

During the evaluation year the external evaluator observed workplace literacy classrooms in action at several of the manufacturing companies described earlier in this report. He also conducted extensive discussions with the Project Director and teaching staff, and with supervisors and employees at several of the companies.

Conclusions: Based on the foregoing activities and the data presented above, certain conclusions regarding the workplace literacy project under review seem appropriate:

(1). THE CENTER/CCSD #54, Management Association of Illinois (MAI) and the ten manufacturing companies involved in the project formed successful partnerships to bring workplace literacy programs to 948 employees in the Chicago area. Although 108 courses were provided (108% of goal), the project served 948 workers which constituted 62% of the total originally anticipated in the proposal to the U. S. Department of Education.

(2). The Project Director and staff demonstrated that they have developed interpersonal skills and operational procedures that permit them to repeatedly enter into a business, set-up an education coordination team, conduct a basic skills needs analysis and assessment with managers, union members and employees, develop job-related assessment instruments and administer them, develop and deliver job-related English language, reading/writing, and mathematics programs on company sites at times convenient to the employers and employees.

(3). Supervisor judgments, job-related test score data, and employee judgments all converge to suggest that the workplace literacy programs (a) produced improvements in job-related basic skills; (b) in many cases improved productivity through the reduction of wastage and errors; (c) improved morale and employee confidence on the job, at home, and in the community and (d) contributed not only to the organizational effectiveness of the companies involved but also to the achievement of National Education Goals 1 and 6 in the Goals 2000 Act.

Recommendations: The recommendations have to do with actions to increase the amount of usable data in future projects.

(1). The external evaluator should be involved earlier in the project. This could result in the development of assessment instruments earlier and in their earlier use to obtain a larger corpus of information that is more representative of the total number of courses offered and employees served.

(2). THE CENTER has now conducted work with over forty different companies in the Chicago area. It should now be possible to draw upon the body of job-related materials and tasks from previous projects to develop alternative forms of job-related assessments that sample across various specific jobs, are normed on regional workers and which could be used as pre-and post-tests in each new program to determine the extent to which the workplace literacy training results in more generalizable work-related basic skills. This could be done with consultation from psychometricians in the Chicago area.

(3). Consideration should be given to the use of a brief, 20 minutes or so, assessment instrument that provides an indication of how well employees perform relative to a national sample. Something like the TABE locator test, or a quick test of vocabulary that provides national percentiles would be useful to indicate the degree of literacy development is needed

to achieve high levels and how much is actually achieved in these brief workplace literacy programs.

(4). Future projects should consider the various organizational functions identified in the Organizational Effectiveness interview and how the project can increase the numbers of "yes" judgments. Perhaps an informational brochure and a briefing could be developed that could educate managers and supervisors about the various OE functions and suggest how they could get public relations, recruitment, etc. benefits from participating in the project.

(5). Future projects should consider the various categories of benefits on the Employment Development interview and develop ways to increase benefits. For instance, a simple pamphlet or a video in English, Spanish and other high frequency languages might be developed to explain the national education goals and how the employees can use their workplace literacy experience to contribute to the various goals.

Chapter 4

Testing and Accountability in Adult Literacy Programs in the Workforce Education Act of 1998

A need for better standards and indicators for accountability in federal adult literacy programs was codified in the Government Performance and Results Act (GPRA) of 1993. In September of 1995, a General Accounting Office report entitled *Adult Education: Measuring Program Results Has Been Challenging (GAO/HEHS-95-153)* was released. The GAO study of the federally and state-sponsored adult literacy education system indicated that progress in achieving GPRA in the federal adult education program had been stymied because "...program objectives have not been clearly defined and questions exist about the validity and appropriateness of student assessments and the usefulness of nationally reported data on results "(p.23).

In June of 1997, the GAO produced another report entitled *The Government Performance and Results Act: 1997 Government wide Implementation Will be Uneven (GAO/GGD-97-109.* This report found mixed results in performance accountability across government agencies and observed that among the significant challenges many agencies face are those that "...involve methodological difficulties in identifying performance measures or the lack of data needed to establish goals and assess performance." (p. 6).

To facilitate the accountability of the federal adult education program, Congress passed the new Workforce Investment Act of 1998 with Title II, The Adult Education and Family Literacy Act. Title II calls for states to develop five year plans that include, among other things, performance measures described in section 212 of the Adult Education and Family Literacy Act. Section 212 requires "core indicators" of performance that include:

- Demonstrated improvements in literacy skill levels in reading, writing and speaking the English language, numeracy, problem-solving, English language acquisition, and other literacy skills.

- Placement in, retention in, or completion of, post-secondary education, training, unsubsidized employment or career advancement.

- Receipt of a High School diploma or its recognized equivalent.

The Adult Education and Family Literacy Act also requires that levels of performance for each indicator be established, and that the levels "...be expressed in an objective, quantifiable, and measurable form; and ... show the progress of the eligible agency toward continuously improving in performance." This state and local information is to be used by the U. S. Department of Education (USDOE), Office of Adult and Vocational Education (OVAE), Division of Adult Education and Literacy (DAEL) to report its progress in meeting the accountability standards of the Government Performance and Results Act of 1993.

This trend to continue to seek more effective methods for accountability in government programs, including adult education will likely be a hallmark of federal activities well into the first decade of the 2000s. For this reason, the present chapter provides information that can be helpful to practitioners in selecting and using standardized tests as "core indicators" of learning in adult literacy programs, whether in the workplace or elsewhere. The

discussion of concepts, issues, and definitions may help program administrators and teachers to more wisely use standardized tests and alternative assessment methods for program evaluation. To this end, topics such as *reliability* and *validity* are discussed in the context of specific problems providers frequently face, rather than as separate psychometric concepts.

Overview. The chapter first addresses aspects of the earlier Adult Education Act of 1988 that address the definitions of standardized tests used by the federal government. This provides insights into the thinking of federal officials regarding standardized tests, and it reveals some of the issues surrounding the uses of standardized testing in adult workforce education. Additionally, it calls attention to technical terminology and other aspects of standardized testing that may be unfamiliar to many who are presently or about to be involved in ABE or ESL program development and implementation.

Next, the nature and uses of standardized tests are discussed. The purpose is to elaborate on the federal definition and discussion, so that users of standardized tests in adult education programs will have a better understanding of what standardized tests are and how to use them appropriately. This section answers questions such as, What does it mean to say that a test is standardized? What is a norm-referenced test? What is a criterion-referenced test? What is competency-based education and how does it relate to the use of norm- or criterion-referenced tests? What is a curriculum-based test?

The nature and uses of standardized tests is followed by discussion of special topics in the use of standardized tests, including: What to do about "negative gain" scores, that is, when students do poorer at the end of the program than they did at the beginning? What is the difference between "general" and "specific" literacy and when should programs assess each? What is predictive validity and what does it have to do with assessment in ABE and ESL programs? How does a test that is developed using item response theory differ from traditional tests? What are some special problems in testing in ESL programs? What are "alternative assessment" methods? What kind of assessment system can be developed to meet instructional purposes and State and federal requirements for accountability?

FEDERAL INTERESTS IN STANDARDIZED TESTING IN ADULT EDUCATION

Prior to the Workforce Investment Act of 1998, the Adult Education Act, as amended in 1988, required State adult education agencies to "gather and analyze data (including standardized test data) to determine the extent to which the adult programs are achieving the goals set forth in the [State] plan..."[1]

In implementing the Adult Education Act, the U. S. Department of Education Rules and Regulations for evaluating federally supported State Adult Education Programs required that State Education Agencies "gather and analyze data on the effectiveness of all State-administered adult programs, services, and activities - including standardized test data...."[2]

The U. S. Department of Education offered a definition of a "standardized test:"

> A test is standardized if it is based on a systematic sampling of behavior, has data on reliability and validity, is administered and scored according to specific instructions, and is widely used. A standardized test may be norm-referenced or criterion-based. The tests may, but need not, relate to readability levels, grade level equivalencies, or competency-based measurements.[2]

For many adult educators, concepts such as "standardized," "norm-referenced," criterion-referenced," and other concepts related to standardized testing may be little understood. These and other concepts related to testing are discussed next to provide adult educators with a better basis for making choices in response to State and federal evaluation and accountability requirements that performance levels "...be expressed in an objective, quantifiable, and measurable form."

NATURE AND USE OF STANDARDIZED TESTS

As noted above, a *standardized* test is a test that is administered under *standard conditions* to obtain a sample of learner behavior that can be used to make inferences about the learner's ability. A standardized test differs from an *informal* test in that the latter does not follow a fixed set of conditions. For instance, in a standardized reading test, the same reading materials are read by different learners following the same procedures, answering the same types of questions and observing the same time limits. The purpose of the standard conditions is to try to hold constant all factors other than the ability under study so that the inference drawn about that ability is *valid*, that is, true or correct.

Standardized tests are particularly useful for making comparisons. They let us compare a person's ability at one time to that person's ability at a second time, as in pre-and post-testing. They also permit comparisons among programs. However, for the tests to give valid results for making such comparisons, they must be administered according to the standard conditions.

By understanding the logic of *standardization* in testing, programs can strive to keep the conditions of test administration from affecting test performance. Here are some *things to avoid*:

Avoid: *Ignoring time standards*. Here is a simple illustration of the reasoning behind the methodology of standard conditions. If a program wanted to compare a group of learners' post-program reading ability to their pre-program ability, and it only gave them *fifteen* minutes to complete a hundred items on the pre-test, then it would not be appropriate to let them have *thirty* minutes to complete a comparable set of items at the post-test. Using such different conditions of test administration, one could not infer that the learners' greater post-test scores indicated a true gain in ability over the pre-test scores. It might simply indicate that the learners were able to complete more items because there was more time. In this case, then, the learners' abilities had not increased. Rather, the conditions under which the test was administered were changed. They were not standard for both the pre- and the post-tests. And these changed conditions of administration may have produced the observed increase in test scores.

Avoid: *Testing the first time students show up for a program.* Many adult students will not be very comfortable at the first meeting. They may be nervous and frightened about taking a test. They may also be unprepared in test-taking strategies. Because of this psychological condition of the learner, they do not meet the conditions of standardization of most tests, which assume a more-or-less relaxed, test-experienced learner. If pre-tested under their first meeting psychological conditions, learners' true abilities may be greatly underestimated. Then, at the post-test, after they have had time to adjust to the program, its staff, and have had practice in answering test questions similar to the standardized tests, their post-test scores may be higher. But in this case, much of the gain may represent the change in the learners' emotional conditions, and not gain in the cognitive ability (e.g., reading, writing, mathematics) that is the object of assessment.

The increase in post-test scores over pre-test scores due to the kinds of psychological factors discussed are sometimes called "warm-up," "surge" or "practice" effects. Such effects may be particularly troublesome when pre- and post-testing are separated by only a few hours. Some programs may have capitalized on such effects in claiming to make one, two or more "years" gain in reading or mathematics in just 15 or 20 hours of instruction. In general, pre-testing should not be accomplished until learners have had an opportunity to adjust to the program and practice their test-taking skills.

Types of Standardized Tests

Scores on standardized tests do not have much meaning in and of themselves. If a learner correctly answers 60 percent of items on some standardized test, it is not clear what that means in the absence of other information that helps us *interpret* the score. We do not know if 60 percent indicates high ability or low ability in the domain being assessed (for example, reading). For instance, if every other adult similar to the learner scores 90 percent correct, then we would probably conclude that 60 percent was an indicator of low ability. To interpret the score, we need other information to which the observed score can be *referenced* or *based,* that is, compared and related.

The federal definition given above notes that standardized tests may be norm-referenced, criterion-based, or competency-based. But it is not always clear just what different scholars or practitioners mean by these terms. The following discussion is meant to provide a common frame of reference for program operators for understanding the various types of standardized tests that are available.

Norm-Referenced Tests. All human cognitive ability is socially derived. That is, the language one uses, the concepts used for thinking and communicating, the logic of reasoning, the types of symbols and symbolic tools (e.g., tables, graphs, figures, bus schedules, tax forms, etc.), and the bodies of knowledge stored in people's brains or in books are developed by individuals being reared in social groups.

Because of the social basis of cognition, many standardized tests have been developed to permit a learner's score to be interpreted in relation to, or, stated otherwise, in *reference* to the scores of other people who have taken the test. In this case, then, an individual's standardized test score is interpreted by comparing it to how well the referenced group *normally* performs on the test. If the individual learner scores above the average or norm of the referencing or *norming* group, the person is said to be above average in the ability of interest. If the learner scores below the average of the referencing group, he or she is said to be below average in the ability.

Grade level norms. In adult literacy education programs, standardized tests are frequently used that have been normed on children in the elementary, middle, and secondary school grades. In this case then, the adult learner's score on the test may be interpreted in reference to the average performance of children at each grade level. If an adult's score on a reading test normed on grade school children is the same as that of a child in the eighth month of the fourth grade, the adult would be assigned an ability level of 4.8. If the adult's score was the same as the average for school children in the sixth month of the ninth grade, the adult would be said to be reading at the 9.6 grade level.

Interpreting these grade level scores for adult learners is not straightforward. For instance, the score of 4.8 does not mean literally that the adult reads like the average child in the eighth month of the fourth grade. In fact, in one research study adults reading at the fifth grade level were not as competent at other reading tasks as typical fifth grade children

(Sticht, 1982). This is not too surprising when it is considered that the child is reading at a level that *defines* what is *typical* for the fourth grader, while the adult in our relatively well-educated and literate society who reads at the fourth grade level is well below the average for adults.

What the fourth grade score for the adult means is that the adult reads very poorly relative to other adults who may score at the ninth, tenth, or twelfth grade levels on the test. While the grade level score is based on the performance of children in the school grades, the interpretation of the score should be based on the performance of adults on the test. For this reason, standardized tests such as the Tests of Adult Basic Education (TABE) or Adult Basic Learning Examination (ABLE) provide norms for adults in adult basic education programs and other settings that permit test users to interpret scores both in grade levels (grade-school referenced norms) and in relation to adult performance on the tests.

Identifying differences among readers. The major use of norm-referenced test scores is to identify differences among a group of people for some purpose. The norm-referenced tests indicate how people perform relative to the norming group. For instance, are they below or above the average of the norming group.

The most widely used standardized, basic skills (reading, mathematics) test that is normed on a nationally representative sample of young adults (18 to 23 years of age) is the Armed Forces Qualification Test (AFQT). This test has been specially designed to permit the armed forces to rank order young adults from those very low to those very high in basic skills and to screen out the least skilled from military service. The U. S. Congress has passed a law prohibiting young adults who score below the tenth percentile on the AFQT from entering military service.

Adult education programs frequently use norm-referenced reading tests to identify those with reading scores below the fourth or fifth grade levels, those scoring between the fifth and ninth grade levels, and those scoring at or above the ninth grade level. These categories are frequently used to assign adults to different levels of reading instruction: basic or beginning reading, mid-level reading, and high school equivalency (General Educational Development - GED) education.

The use of standardized, norm-referenced tests for selection or placement is not an altogether accurate procedure, if for no other reason than the fact that no test is perfectly *reliable*. That is, because of the differences in people's psychological conditions from time to time, and variations in the physical conditions of testing (for example, it may be very cold, or too hot, or too noisy one day, and so forth), people do not usually score the same on tests from one time to the next.

Also, when multiple-choice tests are used that have been designed to discriminate among a wide-range of ability levels, the tests will contain some very easy items, some average difficulty items, and some very difficult items. The multiple-choice format permits guessing. These conditions mean that a person may score correctly on some items by chance alone on one day, but not the next. This produces artifacts that should be avoided in adult education program evaluation.

Avoid: *Regression to the mean.* Because of the imperfect reliability of tests as discussed above, a phenomenon that has plagued adult education programs for decades is regression to the mean. This usually happens when a group of adults is administered as a pre-test, a standardized test that has been normed using traditional test development methods, and a part of the group is identified as low in ability and sent to a program. Then, later on, when just the low group is post-tested, it is found that the average post-test score

is higher than the pre-test score. Under these circumstances, the program offers the gain between pre and post-test scores as evidence of the effectiveness of the program in bringing about achievement.

However, regression to the mean is a statistical process that generally operates under the foregoing conditions. Whenever a low-scoring group is separated off from the total group and then retested, the average score of the post-test will generally be larger than the average score of the pre-test. This is due to the fact that many people are in the low group on the pre-test because they guessed poorly or did not perform well due to anxiety, lack of recent practice in test-taking and so forth, as mentioned earlier. So, when they are retested, their average score moves up toward (that is, regresses toward) the mean (or average) score of the total group on which the test was normed.[3]

Such warm-up and regression effects can be quite large. In one study, military recruits new to the service were tested with a standardized, grade-school normed reading test. Those scoring below the sixth grade level were retested two weeks later, with no intervening reading instruction, and those who scored above the sixth grade were excluded from the study. Two weeks later, the remaining recruits who scored below the sixth grade level were retested with a third form of the reading test, and those who scored above the sixth grade level were excluded. This process reduced the number of people reading below the sixth grade level by 40 percent (Sticht, 1975)!

Regression effects can be reduced in several ways. One is to use the retesting procedure discussed above. Obviously, this requires quite a commitment to testing. It also requires the use of standardized tests with at least three comparable forms, one for the first testing, a second for the next testing of the group identified as low on the first testing, and a third for the post-testing of the group identified in the second testing who were placed in the program of interest.

Regression effects can also be reduced by not testing learners until they have adjusted to the program and obtained some practice in test-taking as noted earlier.

In another approach to managing regression effects, scores on post-tests may be adjusted for regression by using the correlation between pre and post-test scores. This permits the prediction of post-test scores from pre-test scores. Then, actual post-test scores can be compared to the predicted scores. Only the gain that exceeds the predicted post-test scores is then used to indicate program effectiveness. This procedure requires technical assistance from a knowledgeable statistician or psychometrician.

Regression effects may also be estimated and adjusted for by comparing the program group to a group with similar pre-test scores which does not receive the educational program being evaluated (though note that the control group should receive some practice in test-taking, to offset the "warm-up," "surge" or "practice" effects discussed above). This "treatment" and "no treatment" groups comparison permits programs to adjust their gains for regression.

Use of tests with very low probabilities for guessing can also reduce regression. This will be discussed later on in regard to the problem of "negative gain."

Criterion-Referenced Tests. The concept of criterion-referenced assessment was stated in contemporary form by Glaser and Klaus (1962). The concept was advanced as a contrast to the wide-spread method of grading in educational programs known as grading "on the curve." In grading based "on the curve," learners' grades depend on how well everyone in

the class or other norming group performs. An individual learner's grade is determined in relation to the grades of others. Therefore, if everyone in the class performs poorly, a low mark, say 60 percent correct, may be assigned a relatively high grade, say, a "B." Yet, if everyone performed well, a mark of 60 percent correct might be assigned a grade of "D."

In criterion-referenced testing, an absolute standard or criterion of performance is set, and everyone's score is established in relation to that standard. Thus, 90 percent correct and above might be necessary to receive a grade of "A," 80 to 89 percent correct for a "B," and so forth. In criterion-referenced testing then, learners' achievement in an instructional program is assessed in terms of how well they achieve some absolute standard, or criterion of learning, rather than by comparison to a norming group.

Using a norm-referenced test is like grading "on the curve." If the norming group improves overall, then tests may be renormed to adjust the average score higher. There will always be somebody below average. This does not permit one to say, then, how well someone has or has not mastered some body, or as it is called in test development, some *domain* of knowledge or skill.

Criterion-referenced testing had its roots in the behavioral psychology of the 1950's and 1960's, and was closely related to the development of self-paced, individualized, more-or-less carefully pre-programmed instruction. In instructional programs following this approach, a domain of knowledge and skill is carefully defined. Learning objectives that can be assessed are specified, and units of instruction, frequently called "modules" are developed to teach the various subsets of knowledge and skill identified by the learning objectives.

With the modules in place, learners are introduced to a module preceded by a pre-module test, to see if they already know the material to some pre-determined criterion, e.g., 90 percent correct. If the learners pass the pre-module test, they go on to the next module with its pre-module test and so forth. If a pre-module test is failed, then the learner is assigned the study materials and lessons of the module in question, and then is administered a post-module test to see if he or she can perform at the desired criterion.

In this criterion-referenced approach to assessment, learner gain is interpreted in terms of how many units of instruction are mastered at the prescribed criterion level and not in terms of the learner's change relative to a norming group.

Competency-Based Education and Testing. Closely related to the concept of criterion-referenced testing is the concept of "competency-based" education. Just as criterion-referenced testing was put forth in opposition to the practice of grading "on the curve," a practice which obscures just how much learning may take place in a program, the concept of competency-based education was put forth in opposition to the traditional practice of awarding educational credit or certification on the basis of hours of instruction or number of courses completed. Such factors do not reveal the actual competence developed in the program of instruction.

The major factor distinguishing "competency-based" from "traditional" education is the idea that a learner's progress in the course should be based on the demonstration that new competence has been achieved, not on the basis of the number of hours or courses in which the learner has participated.

Because competency-based programs typically identify learning objectives very specifically, they tend to use criterion-referenced assessment. Sometimes, both criterion- and norm-referenced tests are used in competency-based programs. For instance, in the

Job Corps program, or its "civilian" adaptation, the Comprehensive Competencies Program (CCP), a norm-referenced test, such as the TABE, is administered as a pre-test to determine the learner's general level of skill for placement into the instructional modules of the program. Then criterion-referenced assessment is used to indicate whether or not learners are mastering the specific course competencies, as in the pre- and post-module assessments mentioned above. Finally, norm-referenced, post-course tests are used to indicate growth in the "general" ability to which the specific competencies contribute (Taggart, 1985).

What makes the course "competency-based" is the fact that criterion levels of achievement on the norm-referenced tests are established, such as achievement of the 8th grade level, before promotion is made to the next level of education, such as high school equivalency instruction. The 8th grade level of achievement is the criterion that must be achieved for promotion to the next level of instruction. As this illustrates, norm-referenced tests may be used as criterion-referenced tests in competency-based instruction.

In the Comprehensive Adult Student Assessment System (CASAS) hundreds of basic skills (listening; reading; mathematics) competencies judged to be important to be mastered by adult basic education learners have been identified. For each of the hundreds of competencies, a number of test items have been developed to assess mastery of the competencies at different levels of difficulty. These thousands of test items have been formed into a number of standardized tests to determine if adult learners can perform the competencies at deferent levels of ability. Because the test items are based on the competencies identified earlier, the CASAS tests are referred to as competency-based tests (Davis, et. al, 1984).

Curriculum-Based Assessment. Typically, in criterion-referenced or competency-based programs, developers first identify what the important objectives or competencies are that should be learned. Next, test items are developed to determine whether learners already possess the competencies or if instruction is needed to develop certain competencies. Then, various commercially available curriculum materials with a variety of learning exercises are identified that teach each of the competencies so that teachers can select the materials their learners need to master.

This approach, then, is a form of "teaching to the test," even though the exact contents of the assessment instruments may not appear in the curriculum to avoid directly teaching to the specific test items. The competency-based test is used, rather, to indicate the degree of transfer from the curriculum to the application of the new learning.

In curriculum-based assessment decisions are first made about what is important to be taught. Then a curriculum is developed, which may or may not be a formally, pre-developed series of learning experiences. Sometimes, very individualized content and learning activities are improvised by teachers and learners as a dynamic process. Finally, tests are constructed to "test to the teaching." Here the intent is to determine whether what is being taught is being learned and, if not, how instruction should be modified (Bean, et. al, 1988).

In this case then, what is learned becomes the new competence gained in the program. The difference between the competency-based test and the curriculum-based test lies in the direction of test development. In the competency-based programs, the competencies are identified first and the curriculum is designed to help the learner achieve these specific competencies.

In the curriculum-based test, the learner's specific learning activities generate new competence that can then be certified through the development and administration of a curriculum-based test.

The idea of curriculum-based assessment arose from disappointment with the use of nationally standardized tests in which the contents and skills being assessed did not match precisely what was being taught in the schools (Fuchs & Deno, 1981). This results in part from the requirement that, to market a test nationally, test developers cannot tie the test too closely to any particular curriculum. Further, they assess learning that takes place in both school and out-of-school experiences. As a consequence, the tests are generally not sensitive to the specific content (concepts; vocabulary; skills) that is being taught in a particular curriculum.

To appear to be related to all curricula, tests frequently use words that appear precise, but are not. For instance, assessing "Vocabulary Skills," as though "vocabulary" is a generalizable "skill," which it is not, instead of specific knowledge, which it is. In general, "skills"-oriented terminology is used to suggest that "process" ability and not content knowledge is being assessed. But this ignores that fact that all "process" requires some content on which to operate.

For workplace basic education programs, in which there is generally precious little time for adults to participate, the "skills" focus is recognized as not being sensitive to the particular job-linked content that is taught. To a large extent, that is why there is very little increase in the standardized test scores of most adult learners in the relatively brief time that they attend programs. The nationally standardized and normed tests are not sensitive enough to the specifics of what is being taught in the program. Among others reasons, this is why many programs are searching for alternatives to such standardized tests. There is a desire for more curriculum-based assessment so that learners' "true" gains can be detected. This is discussed further under the topic of alternative assessment, below.

SPECIAL TOPICS IN THE USE OF STANDARDIZED TESTS

Certain questions about the uses of standardized tests and alternative assessment methods that policymakers, administrators, teachers, and evaluators have raised from time to time are discussed below. These include:

What to do about "negative gain" scores, that is, when students do poorer at the end of the program than they did at the beginning?

What is the difference between "general" and "specific" literacy and when should programs assess each?

What is "item response theory" and what does it imply for testing in ABE and ESL programs.

What is predictive validity and what does it have to do with assessment in ABE and ESL programs?

What are some special problems in testing in ESL programs?

What are "alternative assessment" methods and what are their advantages and disadvantages?

What kind of assessment system can be developed to meet instructional purposes and State and federal requirements for accountability?

Negative Gain

In ABE or ESL programs it is not unusual to find that 10-20 percent of learners score poorer on the post-test than they do on the pre-test. Therefore, when the post-test score is subtracted from the pre-test score to calculate the gain score, the gain is a negative number (Taggart, 1985; Caylor & Sticht, 1974).

It is possible (though not very probable, perhaps) that negative gain may occur because learners on the pre-test do not work at any given item too long, because they think they cannot perform the test task, and so they simply guess at all the items. On the post-test they spend more time on each item because they have new competence and think they should not guess but try to actually comprehend and perform each item. This could lead to more accurate, but fewer test items being completed at the post-test, and hence a negative gain score.

Generally, however, negative gain reflects guessing or other regression effects. In this case, guessing on the pre-test is better than guessing on the post-test and this leads to negative gain. This can be reduced by using tests that require constructed responses, or that offer many alternatives for multiple choice tests. The latter reduces the effects of guessing. In one study where tests with very low probability for guessing were introduced, negative gain was reduced from 30 percent to 6 percent (Sticht, 1975).

For those programs in which tests with higher potential for negative gain exists, and this includes all multiple choice tests, frequency distributions showing numbers and percentages of learners making various amounts of negative and zero gain should be included. This permits evaluators to gauge the amount of regression occuring in the program. Simply showing average pre-and post-test scores that includes the zero and negative gains obscures this valuable information and produces inaccurate indications of lower improvement in the program than actually occurs.

"General" and "Specific" Literacy

Learner-centered literacy instruction in which the functional context of the learner dictates the curriculum differs from literacy education based on the idea that adult basic education should replicate the school grades and eventually lead to a high school equivalency certificate. Literacy education aimed at giving the adult learner the same kinds of knowledge and information processing abilities as possessed by typical high school graduates is known as "general" literacy.

Literacy education aimed at providing adult learners with some particular, more circumscribed body of knowledge and information processing abilities, such as those involved in a particular line of work (e.g., automobile mechanic), life role (e.g., parent) or life activity (e.g., reading tax manuals) is known as "specific" literacy.

For many reasons, adult learners do not always have a lot of time to spend in a basic skills program. For instance, if they are unemployed and need to learn a job quickly, then time in a general literacy program that aims to recapitulate the public school curriculum will prolong the adult's entry into job training and hence into gainful employment. Furthermore, evidence suggests that "general" literacy education does not transfer much to improve "specific" literacy in the relatively brief (25,50,100) hours of education that adult learners will choose to attend. However, "specific" literacy training may produce as much

improvement in "general" literacy as do typical "general" literacy programs (Sticht, 1975; Sticht, 1988).

For these reasons, workplace literacy programs generally integrate basic skills training with job knowledge and skills development. For instance, a person desiring to learn to be an automobile mechanic is given reading, writing, and mathematics education using automobile mechanics training textbooks or technical manuals and performing functionally relevant, literacy task performance.

Following similar reasoning, if learners wish to read books to their children, literacy providers can teach "specific" literacy by teaching learners about children's books, how to read and interpret them with their children, and so forth. Or, adults desiring to read a tax manual can be taught literacy using a tax manual and special materials to develop "specific" ability in reading tax manuals.

A very large amount of materials and procedures exist for teaching English for Specific Purposes (ESP) in English as a Foreign Language or in English as a Second Language (ESL) programs. Such ESL programs are sometimes known as VESL-Vocational English as a Second Language- programs.

In all these specific literacy or language programs, assessment instruments can be developed that are curriculum-based, as discussed above. These "specific literacy tests" will be most sensitive to the adult learners' goals and gains. Programs can also use "general literacy" tests to indicate the degree of generalizability that occurs in the "specific" literacy program.

Item Response Theory (IRT)

With the growth in use of tests such as the Comprehensive Adult Student Assessment System (CASAS) (Davis, et. al, 1984) and the National Adult Literacy Survey (NALS) (Kirsch, Jungeblut, Jenkins, & Kolstad, 1993) more ABE and ESL program providers are reading about *item response theory*.

The CASAS and NALS (as well as the International Adult Literacy (IALS) and several other tests widely used in adult basic education programs) have been developed using newer psychometric methods based on item response theory. In general, IRT is a method for scaling individual test items for difficulty in such a way that the item has a known probability of being correctly completed by an adult of a given ability level.[4] For instance, on the CASAS scale, an adult learner with an ability score of 215 has a fifty percent chance of passing all items that are in the item bank that are also scaled at 215. For items rated below 215, the learner has a greater than fifty percent chance of getting the items correct, and with items above 215 the learner has less than a fifty percent chance of getting the items correct.

If a program has a test item bank of several thousand items that are all on the same IRT scale, it is possible to administer a relatively small sample of the items in a test and from this small sample of items, know the probability that the learner can perform each of the other items in the bank. Obviously this is useful for diagnosing particular competencies that a learner may need to develop further.

Traditionally developed tests do not provide probability of performance estimates for items not in the test. Furthermore, traditionally developed, norm-referenced tests have to be renormed every time the items in the test are changed. But with an IRT-based test, items

from a bank can be reconfigured into different forms of tests without having to renorm the test. This means that it is easier for programs to tailor tests for their particular curriculum and for learner needs.

In particular, IRT is useful for developing multiple forms of tests that are suitable for a restricted range of ability. This permits more reliable estimation of ability for learners within the range being studied.

Though the power of IRT will ensure that most future test development will utilize this psychometric technology, it should be noted that there is nothing in the IRT that ensures the *validity* of the tests. Validity refers to whether or not a test actually measures what it purports to measure, *and nothing else*.

But absolute validity is a very difficult thing to achieve. All paragraph reading comprehension tests, for instance, measure not only skill in decoding printed language and performing tasks such as identifying the main idea, but also a learner's background knowledge related to what is being read. This is true regardless of whether the tests are developed using traditional or item response theory psychometrics.

Predictive Validity

In the discussion of Item Response Theory, *validity* was defined as referring to whether or not a test measures what it purports to measure and only that.

There is, however, another type of validity that is assuming greater importance in ABE and ESL. This type of validity is called *predictive validity* . Predictive validity refers to how valid or accurate a test is for predicting some future behavior of learners. It is growing in importance as such federal programs as the Workforce Investment Act of 1998 focus on improving basic skills to the levels needed for performing successfully in job training or on the job. In work-oriented literacy programs, the focus is on identifying participants whose basic skills are judged to be (i.e., predicted to be) too low for employment. Under such programs, adults identified as "functionally illiterate" may be denied job training because of their low levels of basic skills. They may be required, instead, to participate in basic skills courses to qualify for job training or to continue to receive their welfare benefits, or both.

Predictive validity is also important in pre-GED testing to determine whether learners qualify to attempt the GED high school equivalency examination. For instance, the CASAS scales suggest that learners with scores of 224 or below are functioning below a high school level, while those with scores at or above 225 can profit from instruction in GED preparation (The CASAS System, 1989). The Official GED Practice Tests are used "...to provide *general* indications of readiness to take the full-length GED Tests (American Council on Education, 1989)."

All uses of basic skills tests to indicate "readiness," ability to "profit from instruction" and that prevent learners from entering into some desired job or job training program are predicting that learners who score below a certain level on the basic skills test will not be successful in the future activity for which the basic skills test serves as a screen. The question for predictive validity is, does the test score criterion accurately (that is, validly) predict who will and will not be able to perform satisfactorily in the job, job training, or GED test-taking situation?

In studies of the predictive validity of the most widely used basic skills test, the Armed Forces Qualification Test (AFQT), it was found that of those that military selection policies had predicted to fail in job training and on the job, eight out of ten actually performed

satisfactorily (Sticht, Armstrong, Hickey, & Caylor, 1987). These data, from an organization that has studied this type of assessment for seventy years at a cost of at least $500 million, should caution the "gatekeeping" use of basic skills tests in workfare/welfare, workplace literacy, and JTPA programs.

No major gatekeeping decision should be based solely on the results of a single standardized test score. Adult education providers should use interviews, past employment experiences, and work sample procedures to counsel learners about their probabilities of success in future activities beyond the boundaries of the basic skills program.

There are well-established laws, and many precedent-setting legal cases to establish a basis for adult learners to challenge test use that adversely impacts them by delaying or preventing access to gainful employment (Gifford, 1989). To date, no studies have been found of the predictive validity of standardized tests used in workfare/welfare basic skills programs, workplace literacy programs or GED preparation programs.

English as a Second Language (ESL)

A growing share of adult basic education is concerned with English as a Second Language programs. In 1991-92, ESL participants made-up 42 percent of students in adult education (U. S. Department of Education, 1993). In California, ESL learners make-up close to 80 percent of participants in ABE (Dixon, Vargo, & Campbell, 1987).

Using standardized tests with ESL learners incorporates all of the problems discussed earlier in this report. Additionally, however, special difficulties are encountered because of the wide differences in the language, cultural, and educational backgrounds of the ESL learners.

For instance, many ESL learners come from groups for which there is no written language (e.g., Hmong, Mien) and so it cannot be assumed that they have general, "world" knowledge of the forms and uses of written language (Savage, 1983). Others, however, may be highly educated and literate in their native language, but simply unable to speak and comprehend English. Given this large range of differences among ESL learners, there is a need to determine, through interviews with learners or their associates, the non-English language education and literacy status of ESL learners prior to administering assessment instruments.

The major difference between ABE and ESL students, of course, is their knowledge of the English language. Most adults, even the highly literate and educated, are reticent about speaking a foreign language. ESL learners are no different from other adults in this regard. Hence, it is necessary to have a period of adjustment during which learners can develop confidence before proceeding with a formal assessment using standardized tests that require learners to speak. This is similar to the need for a "warm-up" period discussed above.

Because speech disappears as it is produced, the evaluation of English speaking, comprehension, and communicative functioning ability (e.g, knowledge of forms of speech for particular occasions) in a dynamic interaction is difficult. This may lead to test situations in which the types of tasks called for are designed to permit special judgments for ease of scoring to be arrived at, but which also appear "unreal" to both teachers and learners. For instance, standardized tests may not permit normal conversational patterns, questioning of meanings by learners, and sharing of information to accomplish a real-life task (Tirone, 1988). This may lead to an underestimate of the learner's communicative competence.

Generally, in testing in ESL programs, as in other ABE programs, it may be desirable to separate testing for program accountability from testing for instructional decision making.

Alternative Assessment Methods

Problems involved in obtaining valid measures of learners' development in adult literacy programs have stimulated a growing interest in alternatives to standardized tests for assessing learner's progress in instructional programs. The September 1989 issue of *Information Update*, the newsletter of the Literacy Assistance Center, Inc. in New York focusses on alternative assessment methods. The issue provides a good example of the types of problems that program providers experience with standardized tests, and presents a rationale for the need for improved assessment methods.

The major problem addressed by the alternative assessment movement is similar to that discussed under curriculum-based assessment, namely the incongruence between what programs teach, what learner's learn, and what the nationally standardized tests assess. Many of the programs that are experimenting with alternative assessment methods do not follow a prescribed curriculum. Rather, they follow an approach in which a learner's expressed needs form the basis for instruction. This approach is frequently called a *learner-centered* or *participatory* approach, because the learner participates in determining the instruction (Lytle, Belzer, Schultz, & Vannozzi, 1989).

Alternatives to nationally standardized testing include intake and progress *interviews* that record such information as the type of reading the learner does, how much reading in different domains (job, home, community) is accomplished, self-evaluations of reading ability, and judgments of abilities by teachers in staff meetings. The California Adult Learner Progress Evaluation Process (CALPEP) illustrates the interview approach to assessment (Solorzano, 1989).

A second method of alternative assessment is *portfolio* development and evaluation.[5] This is a method similar to that followed by artists, designers, models, writers and others in creative fields of endeavor. Using this method, learners develop portfolios of their work in reading, writing, and mathematics, including both in-class and out-of-class work. Peers, teachers, and learners meet periodically to discuss the learner's work and how it is progressing.

Through these meetings, learners' progress is assessed in areas such as *metacognitive processes* (thinking about, evaluating, and planning their work), *cognitive development* (vocabulary, concept knowledge, and reasoning processes typical of an area chosen by the learner; knowledge of the functions and structure of various types of texts -notes, letters, reports from school, work materials, etc.), and *affective development* (self-understanding and esteem, value of literacy for self, children, and others).

Sometimes direct indicators of competence and its change are obtained by having learners perform, much as a performing artist would. For instance, in a reading program the performance might consist of *reading aloud* (Bean, Byra, Johnson, & Lane, 1988; Hill, 1989). As the learner performs, the teacher may record the oral reading and then later listen to the recording with the learner. Together they evaluate the reading performance for progress in pronunciation, accuracy of word identification, inflection cues to comprehension, and other information identified in participation with the learner.

Assessing Alternative Assessment. There can be no doubting that the alternative assessment methods provide new information about adult learners in ABE, ESL,

workplace, and family literacy programs. Much of this information reflects newer concepts about literacy and other abilities from contemporary cognitive science.

Alternative assessment methods relate very much to the teaching and learning process as it takes place in the classroom in interactions among teachers, learners, peers and the various materials they use and tasks they perform. In general, the richer the descriptive information about these interactions and processes, the more valid will be the understanding of particular programs by both internal (administrators; teachers; learners) and external (local; state; federal) evaluators.

However, while these alternative methods are invaluable for their contributions to learner progress, there are limitations to the exclusive use of such techniques for learner and program evaluation, as those developing these new assessment methods acknowledge (Dick, 1989).

One of the problems identified by alternative assessment providers is the fact that, although standardized, nationally normed tests fail to match program content, administrators, teachers, and millions of other adults can and do perform very well on any or all of the dozens of standardized tests of reading, writing, and arithmetic that are the subject of criticism. The question is raised, therefore, of whether or not adult learners in ABE and ESL programs are being directed to less demanding levels of achievement if they are not evaluated using standardized tests.

It has also been noted that standardized tests

> "...are an integral part of the fabric of our lives. One has to take tests to get into college, to enter the military and to obtain civil service employment, to mention just a few. While such tests should certainly not be the measure of individual student progress in the adult literacy classroom, we ought not ignore the value for students of being familiar with them and being able to use them to their own advantage (Dick, 1989)."

A problem with the sole reliance on alternative assessment methods for program evaluation for public accountability is that nonstandardized methods make it difficult to compare across programs. One goal of the federal guidance on quantifiable and measurable indicators of learning is to make it possible for outside evaluators to know how well one program or group of programs is promoting learning compared to other programs.

Assessing for Instruction and Accountability

Many of the problems with standardized testing experienced by programs are due to the attempt to use one test for both program accountability and instructional decision making. For instance, using the Tests of Applied Literacy Skills (TALS), which is a commercial version of the National Adult Literacy Survey (NALS), for pre and post-testing to report gains in general literacy to state and federal administrators is a program accountability function of the tests.

But using the TALS to assess learning in a specific literacy program, in which learners may choose to read and study a child-rearing manual is an inappropriate use of the test for assessing either instructional needs or progress. In this case, an alternative assessment method is needed, perhaps one in which learners' needs are determined by interviews that include trial readings of manual passages. Then, progress checks using reading aloud and question/discussion periods for checking comprehension might be used to indicate learning in the program.

In one military project, a specific job-related literacy program was developed that used three types of testing (Sticht, 1975). Pre and post-module testing was used in a competency-based, criterion-referenced, testing/teaching curriculum strand. The module tests provided curriculum-based indicators of both instructional needs and progress.

A second testing method was developed in which job-related reading tasks from across six career fields were sampled and included in job-related reading task tests. These tests were used as pre and post-program measures of generalizable growth in work-related (though not job-specific) types of reading skills. They were then normed in grade levels because the military management preferred to indicate program growth in grade levels.

Thirdly, a nationally standardized and normed test was administered pre and post-course to indicate growth in general literacy in grade level units.

As might be expected, in this program, the most learning was indicated by the pre and post-module tests, the next largest increase was in the pre and post-course, work-related tests, and the least increase was in the general literacy tests.

In general, multiple assessments can contribute multiple types of information. Nationally standardized tests, properly administrated, can provide information about broad growth in literacy or mathematics skills. But this growth will typically not exceed one or two "years" in 25, 50 or 100 hours (and this must be obtained with regard to the problems of warm-up and regression discussed earlier). This information can be used for cross-program evaluations of broad ability development.

For instructional decision making, assessment more closely coupled to the curriculum provides the best indicator of what is being achieved by learners in the program. In general, the two important questions here are, "What do learners want to learn?" and "Are they learning it?"

Footnotes

[1] Public Law 100-297, Title III, Part A, Subpart 5, section 352: Evaluation.

[2] Federal Register, August 18, 1989, p. 34435.

[3] Regression to the mean also occurs whenever a high scoring group has been separated from the total group and retested later on. In this case, the average score of the high scoring group will tend to *decrease* as it regresses to the mean of the total group.

[4] More can be learned about Item Response Theory (IRT) in a text and computer assisted instruction program: F. Baker (1985). *The BASICS of item response theory.* Portsmouth, NH: Heinemann Educational Books.

[5] See articles by M. Wolfe and S. Hill in the September 1989 special issue of *Information Update* published by the Literacy Assistance Center, Inc. of New York city. For an earlier application of performance/portfolio-type assessment applied to adult education see R. Nickse (1980). *Assessing life-skills competence.* Belmont, CA: Pitman Learning, Inc.

References

American Council on Education (1989). *The Official GED practice Tests.* Washington, DC: American Council on Education.

Bean, R., Byra, A., Johnson, R. & Lane, S. (1988, July). *Using curriculum-based measures to identify and monitor progress in an adult basic education program. Final Report.* Pittsburgh, PA: University of Pittsburgh, Institute for Practice and Research in Education.

Caylor, J. & Sticht ,T. (1974, April). *The problem of negative gain scores in the evaluation of reading programs.* Chicago, IL: Paper presented at the annual meeting of the American Educational Research Association.

Davis, J., Williams, P., Stiles, R., Wise, J. & Rickard, P. (1984, April). *CASAS: An effective measurement system for life skills.* New Orleans, LA: Paper presented at a meeting of the National Council on Measurement in Education.

Dick, M. (1989, September). From the editor. *Information Update, 6,* pp. 1-2.

Dixon, D., Vargo, M. & Campbell, D. (1987, July). *Illiteracy in California: Needs, services & prospects.* Palo Alto, CA: SRA Associates.

Fuchs, L. & Deno, S. (1981). *The relationship between curriculum-based mastery measures and standardized achievement tests in reading. (Research Report No. 57).* Minneapolis, MN: University of Minnesota, Institute for Research on Learning Disabilities.

Gifford, B. (Ed.) (1989). *Test policy and the politics of opportunity allocation: The workplace and the law.* Boston, MA: Kluver Academic Publishers.

Glaser, R. & Klaus, D. (1962). Proficiency measurement: Assessing human performance. In: R. Gagne (Ed.), *Psychological principles in system development.* New York: Holt, Rinehart, and Winston.

Hill, S. (1989, September). Alternative assessment strategies: Some suggestions for teachers. *Information Update, 6,* pp. 7,9.

Kirsch, I. & Jungeblut, A. (1986). *Literacy: Profiles of America's young adults.* Princeton, NJ: Educational Testing Service.

Kirsch, I., Jungeblut, A., Jenkins, L., & Kolstad, A. (1993, September). *Adult Literacy in America: A First Look at the Results of the National Adult Literacy Survey.* Washington, DC: U. S. Government Printing Office.

Lytle, S., Belzer, A., Schultz, K. & Vannozzi, M. (1989). Learner-centered literacy assessment: An evolving process. In: A. Fingeret & P. Jurmo (Eds.), *Participatory literacy education.* San Francisco, CA: Jossey-Bass.

Savage, K. (undated, circa 1983). *Teaching strategies for developing literacy skills in non-native speakers of English.* Washington, DC: The National Institute of Education.

Solorzano, R. (1989, February). *Analysis of learner progress from the first reporting cycle of the CALPEP field test. A report to the California State Librarian.* Pasadena, CA: Educational Testing Service.

Sticht, T. (1988). Adult literacy education. In: E. Rothkopf (Ed.), *Review of research in education. Volume 15*. Washington, DC: American Educational Research Association.

Sticht, T. (1982). *Evaluation of the "reading potential" concept for marginally literate adults*. Washington, DC: Office of the Assistant Secretary of Defense (Manpower, Reserve Affairs, and Logistics).

Sticht, T. (1975). *A program of Army functional job reading training: Development, implementation, and delivery systems (Final Report)*. HumRRO-FR-WD-(CA)-75-7. Alexandria, VA: Human Resources Research Organization.

Sticht, T., Armstrong, W., Hickey, D., & Caylor, J. (1987). *Cast-off youth: Policy and training methods from the military experience*. New York: Praeger.

Sticht, T. (1985). Understanding readers and their uses of texts. In: T. Duffy & R. Waller (eds.), *Designing usable texts*. New York: Academic Press.

Taggart, R. (1985, March). *The Comprehensive Competencies Program: A summary of 1984 results*. Washington, DC: Remediation and Training Institute.

The CASAS System (1989, November). *GAIN Appraisal Program. III. Third Report*. San Diego, CA: San Diego Community College District Foundation, Comprehensive Adult Student Assessment System.

Tirone, P. (unpublished, circa 1988). *Teaching and testing - can we keep our balance*. New York: Literacy Assistance Center, Inc.

U. S. Department of Education. (1993, September). *National Evaluation of Adult Education Programs. Second Interim Report. Profiles of Client Characteristics*. Washington, DC: Division of Adult Education and Literacy.

Chapter 5

Determining How Many Adults Are Lacking in Workforce Literacy: The National and International Adult Literacy Surveys

If "knowledge-based" nations are to make all of their adults literate enough to compete in the international marketplace, as well as meeting their responsibilities as parents and citizens, how many adults are we talking about? The answer is that it is difficult to say with any degree of certainty. This is because there is not a consensus anywhere on how to define literacy, and all the existing definitions are to some extent arbitrary with respect to how standards of proficiency are set. That is, people are not typically either totally literate or totally illiterate. Rather, they fall somewhere in between. So one of the problems in determining how many adults are likely to be experiencing very difficult times due to their literacy is determining how good is good enough. This problem is illustrated in the context of the 1993 National Adult Literacy Survey (NALS) of the United States[34], modified versions of which were also used in the International Adult Literacy Survey administered in several other industrialized nations[35] (See Table 5.3, below).

The National Adult Literacy Survey (NALS)

In 1993 the National Center for Education Statistics of the U. S. Department of Education reported the results of a survey of the literacy skills of adults aged 16 to over 65 living in households in the United States. Additionally, the survey studied the literacy skills of incarcerated adults.[34] The National Adult Literacy Survey (NALS) used prose, document, and quantitative scales. Literacy scores were reported using scale scores for each of the three different types of literacy task domains. These scale scores ranged from 0 to 500.

Using Item Response Theory (IRT) (see Chapter 4), both people and tasks (items) were given scale scores. For instance, a person with a skill level of 210 would have a probability of .80 of performing a task that has a difficulty level of 210. However, other people with lower skill levels may also be able to perform the task, though with lower probabilities. People with skill levels of 150 have a 32 percent probability of being able to perform a task that is at the 210 difficulty level. People at the 200 level have a 74 percent probability of performing the task. People at the 300 skill level have a 99 percent probability of performing the 210 difficulty level task.

The NALS Literacy Levels

The NALS was the first national survey of adult literacy skills to report data in terms of five levels of skill. The NALS literacy levels are important because they are to be used by the National Governor's Association and the federal government to track the nation's progress on Education Goal Number 6: making all adults literate by the year 2000 [36]. The goal is to get adults to Level 3 in literacy proficiency.

In the NALS, the five levels used to describe categories of proficiency include Level 1 (scale scores from 0 to 225), Level 2 (scale scores from 226 to 275), Level 3 (scale scores 276 to 325), Level 4 (scale scores 326 to 375), and Level 5 (scale scores from 376 to 500). For each of the prose, document, and quantitative scales, all those adults with scores from 0 to 225 were assigned to Level 1, those with scores from 226 to 275 were assigned to

Level 2 and so forth. Table 5.1 shows the percentage of adults assigned to each of the five literacy levels for each of the three literacy scales.

Altogether, the adult population sampled represented approximately 191,000,000 adults. The data in Table 5.1 suggest that some 40 to 44 million adults are in the lowest level of skill, Level 1. Some 50 million are in Level 2, 61 million in Level 3, 28 to 32 million in Level 4 and 6-8 or so million adults are in Level 5.

Table 5.1

Percentage of adults in each of the five NALS skill levels for each literacy scale.

	Level 1	Level 2	Level 3	Level 4	Level 5
Prose	21	27	32	17	3
Document	23	28	31	15	3
Quantitative	22	25	31	17	4
Normal Curve	16	15	38	15	16

For comparison purposes, the percentage of people is given who would fall under the normal or "bell" curve at below -1 standard deviation (S.D.), between -1 to -0.5 S.D., between ± 0.5 S.D., between +0.5 and + 1.0 S.D. and above +1S.D. The data indicate that by using the criterion-referenced standards of the NALS, the percentage of people in the lower two levels is well above what would be expected from a norm-referenced approach in which the mean and S.D. of the population is used to define levels of proficiency. The NALS approach greatly reduces the percentage of those at the highest level (Level 5).

What is The Meaning of the NALS (IALS) Levels?

Being assigned to one of the five levels means that people at the average skill for a given level have an 80 percent probability of being able to perform the average tasks at the given level. For instance, the NALS report indicates that a person with a skill level of 200 would be assigned to Level 1, for which the average task difficulty is about 200 (averaged across the three literacy domains). This means that the person would be expected to be able to respond correctly to 80 percent of the average tasks in Level 1. However, this same person would be expected to be able to correctly respond to over 30 percent of the average tasks at Level 2, about 15 percent of the average tasks at level 3, 8 percent of the average tasks at Level 4 and about 5 percent of the average tasks at Level 5[34, p. 102]. This results from the fact that persons with skill levels below the difficulty level of an item may be able to perform the item correctly, though with a less than 80 percent probability of a correct response.

For example, consider a prose literacy task item that is of 279 difficulty for which a person needs a skill level of 279 to have an 80 percent probability of being able to perform the item. A person with a skill level of 250 has a probability of .62 of being able to perform the item. Because the person has a skill level of 250, on the NALS this would result in the person being assigned to Level 2. This would mean that the person has a .80 probability of being able to perform average Level 2 tasks. But note that the person would also be able to perform Level 3 tasks (which is where a task of 279 difficulty would fall), but not with as high a probability of success. In the NALS report, it is indicated that on either the prose, document or quantitative tasks, a person with a skill level of 250 can be expected to perform 50 out of 100 tasks that are at the average Level 3 task, 25 to 30 percent of the

tasks at Level 4 and 10 to 20 percent of the tasks at Level 5, depending on the type of literacy scale under discussion [34p. 102].

By assigning people to a given skill level, the impression may be formed that the person has *no* ability to perform higher level tasks. But this is wrong. Even though people may be assigned to a lower skill level, this does not mean that they are totally incapable of performing tasks at higher skill levels. In the NALS survey, respondents were asked to rate themselves as to how well they thought they could read and write English. Of those categorized as Level 1 literates, some 66 to 75 percent said they could read and write "well" or "very well." The NALS authors refered to this as the "gap between performance and perception," meaning that the literacy skills of those in Level 1 are low by NALS methods of setting standards for inclusion at one or another level of skill. So the self-perceived skills of the vast majority of those categorized as Level 1 literates, who rated themselves as "well" or "very well" as literates, must be incorrect. They go on to say that "Such a mismatch may well have a significant impact on efforts to provide education and training to adults: Those who do not believe they have a problem will be less likely to seek out such services or less willing to take advantage of services that might be available to them." [34p. 20].

But it is possible that many adults labled as Level 1 literates perceive themselves as quite literate because, as indicated above, they *are* able to perform quite a few tasks at higher levels, even a few at Level 5. It must be kept in mind that simply because people are assigned to a lower level category of literacy level, this does not mean that they are entirely incapable of performing tasks at higher skill levels. They simply do not have a .80 probability of performing higher level tasks. That is, they cannot perform them with the same high level of probability that is required to be categorized at a higher level. This is important to keep in mind when one discusses the numbers of adults in the different skill levels. The numbers can be changed dramatically simply by changing the criterion for being categorized into the different levels.[37] For instance, if instead of requiring that people be able to do 80 percent of the average tasks in a given level, the criterion were changed to being able to do 70 percent of the tasks, then the numbers of people assigned to the lower levels would decrease dramatically.

By using the method of "literacy levels" to categorize people's literacy skills, one may be lead to conclude that people assigned to a given level of skill cannot perform the more demanding types of tasks found at higher levels of skill. Yet that is incorrect and provides an inaccurate indication of the full range of people's literacy skills. Quite possibly, people's perceptions of their literacy ability may be more accurate than the impressions that might be created by the use of the five NALS literacy levels.

Some Major Findings from the NALS

The NALS reported data on the literacy scores of adults across a wide range of age, for persons with special health conditions, for ethnic groups, and for incarcerated populations. Some of the key findings for each of these groups are summarized below.

Literacy and Age. The NALS report indicated that, generally, both education and literacy skills increased for adults from ages 16-18 up to ages 40-54, and then skills dropped rapidly. Adults 55 -64 and those 65 or older performed well below the levels of younger adults, even though their average years of education was not much different from the 16-18 year olds. Summarizing across the three literacy scales, about 44-48 percent of those adults categorized in Level 1 were aged 55 or older, and 32-35 percent were 65 years old or

older. Some 28-32 percent of those in Level 2 were 55 years old or older, and 16-18 percent were 65 or older.

From the NALS data it is not possible to say whether adults' literacy skills rise and then decline or whether the various age groups have performed at the levels indicated throughout adulthood. This would require longitudinal studies. However, the NALS tasks do impose heavier burdens on working memory as they increase in difficulty. In fact, this may be one of the major reasons the tasks increase in difficulty. The authors of the NALS report note that, of several variables that might make tasks more difficult, two of the variables for prose and document tasks are the number of categories or features of information that the reader has to process or match, and the number of categories or features of information in the document that can serve to distract the reader or that may seem plausible but are incorrect. In the quantitative tasks, the number of operations needed to perform the task is given as a factor that may influence the difficulty of the task[34pp. 74, 85, & 94.]

Generally, holding features or categories of information in short term or working memory and then searching through other information places greater demands upon working memory, and there is considerable evidence that working memory performance declines with increasing age[38p. 401]. This may explain, at least in part, the decrease in literacy skills as age increases.

One of the factors that is important for literacy is one's organized bodies of knowledge. The bodies of knowledge are what makes it possible to comprehend printed displays, to reason analogically (i.e., from one body of knowledge to another), and to make inferences (i.e., going from the information given in the display to another body of knowledge in one's mental knowledge base to create yet a third domain of knowledge needed to correctly perform an inference-type task). Generally, these organized bodies of knowledge continue to develop across adulthood and tend to resist deterioration in older age [38p. 401]. While the NALS includes tasks that include knowledge content from health, consumer economics, and others, it does not systematically assess people's organized bodies of knowledge in any domain (e.g., health, science, government, etc.). It is not possible to know whether poorly performing people's primary problems may be their lack of knowledge (e.g., vocabulary, concepts, etc.) or of working memory control, or both. But the rapid decline in performance with ages above 55 suggests a strong component of working memory control in the NALS tasks.

Health Conditions. A major contribution of the NALS was the sampling of adults with various forms of physical, mental or other health conditions. The survey reported that 12 percent of the adult population reported some type of health problem. Significantly, as a type of epidemiological indicator of the self-perceived extent of adult learning difficulties in the U. S. population, some 3 percent (7.5 million) adults reported that they suffered from learning disabilities. Around 60 percent of these adults scored in Level 1, and some 22 percent scored in Level 2. Overall, the average scores for those self-reporting that they had a learning disability were: prose-207; document- 203; and quantitative- 200.

Less than one-half of one-percent reported that they were mentally retarded. Eighty-six to 89 percent of these adults were placed in Level 1, with average scores of: prose-143; document-147; and quantitative- 117.

Race/Ethnicity . The NALS provides the most extensive data on the largest numbers of race-ethnic groups of any previous survey. Table 5.2 shows the percentage of race-ethnic groups falling into each of the five levels of the NALS prose scales. Large percentages (20-89) of Hispanics from the various regions were born outside the United States and

generally had Spanish as their primary language. For the most part, the Hispanic groups with large numbers born outside the United States performed more poorly than Blacks on the literacy scales. Because Hispanics born in the United States are more likely to speak and read English, their scores are higher on the literacy scales. For instance, the Hispanic/Other category includes those who were mostly (68 percent) born in the United States, and their scores are higher than the scores for Blacks. Large percentages (78) of Asian/Pacific Islanders were also born outside the United States. A category of "Other" is also given in the NALS report but is not included in Table 5.2.

Across the age span, Hispanics (grouped together) had fewer years of education (average of 10.2 years) than did Whites (12.8) or Blacks (11.6). Through ages 55-64 Asian/Pacific Islanders had the most years of education (average of 13 years), while among those over age 65, Whites had the most education.

Table 5.2 Percentage of race/ethnic group members in each of the five NALS skill levels for the prose literacy scale.

	Level 1	Level 2	Level 3	Level 4	Level 5	Average Proficiency
White	14	25	36	21	4	286
Black	38	37	21	4	0*	237
Hispanic: Mexicano	54	25	16	5	0*	206
Puerto Rican	47	32	17	3	0*	218
Cuban	53	24	17	6	1	211
Central/So. America	56	22	17	4	0*	207
Hisp. Other	25	27	33	13	2	260
Asian/Pacific Islander	36	25	25	12	2	242
Amer. Indian Alaskan Nat.	25	39	28	7	1	254

* percentages less than 0.5 rounded to zero.

Incarcerated Population. The NALS included a national sample of inmates in federal and state prisons. The sample confirmed what is widely understood in showing that the prison population tends to be quite different demographically than the general adult population. For example, the prison population was mostly males (94 percent), 80 percent were below the age of 40, they were less White (35 percent), more Black (44 percent) and Hispanic (17 percent), and less well educated (49 percent with less than a high school education).

The prison population scored lower on literacy than the general adult population. The average scale scores for the three literacy scales were: prose-246 (272 for the general adult population), document-240 (267 general adult population), and quantitative-236 (271 general adult population). In terms of the NALS literacy levels, looking across the three literacy scales, some 31 to 40 percent of inmates were in Level 1, 32-38 percent in Level 2, 22-26 percent Level 3, 4-6 percent Level 4, and less than 0.5 to 1 percent in Level 5.

Poverty, Income ,Occupational Status, and the Intergenerationl Transfer of Literacy. The NALS confirmed other studies going back over the decades in showing that the less literate are more likely to be found in poverty, on welfare, unemployed or employed in poorly paying jobs, and in the lower status jobs that require less education.

The intergenerational effects of parent's education level on the adult's literacy level was also found in the NALS . Adults whose parents had completed a four year college degree were nine times more likely to have completed a college degree themselves than were adults whose parents had 0-8 years of education (46 percent versus 5 percent).Thirty-two percent of adults whose parents had completed 0-8 years of education had themselves completed only 0-8 years of education, whereas only 5 percent of adults whose parents had completed high school reported that they themselves had completed only 0-8 years of education.

Are the literacy skills of America's adults adequate ?

One of the most important things that the National Adult Literacy Survey (NALS) of 1993 was to do was to provide "...an increased understanding of the skills and knowledge associated with functioning in a technological society."

When the NALS research report directly raised the most important question about literacy and functioning in our technological society, the question that must have motivated the U. S. Congress to ask for the survey in the first place, and the question surely of most interest to corporate America, labor unions, adult educators, and adults themselves, the answer was, at best, disappointing. The report asked, "Are the literacy skills of America's adults adequate? That is, are the distributions of prose, document, and quantitative proficiency observed in this survey adequate to ensure individual opportunities for all adults, to increase worker productivity, or to strengthen America's competitiveness around the world? " 34p. xviii

The NALS authors then went on to answer the question. "Because it is impossible to say precisely what literacy skills are essential for individuals to succeed in this or any other society, the results of the National Adult Literacy Survey provide no firm answers to such questions." 34p. xviii In short, the most important question from a policy point of view was not answered by the NALS (nor has it been answered prior to or since the NALS).

The Arbitrary Nature of Competency Standards. As noted above, in reporting the 1993 National Adult Literacy Survey (NALS) data, the developers assigned adults to five different levels - 1 (low) through 5 (high). To qualify to be at a given level, the arbitrary decsion was made that an adult had to have an 80 percent (p=.80) chance of being able to perform the average task at the given level. Following this decision rule, some 20 percent of adults were placed in Level 1, while 27 percent were placed in Level 2 (prose scale). This led to the quote in many newspapers that "half of America's adults are functionally illiterate!" A sentiment subsequently expressed internationally by leaders in Japan.

But the NALS data also showed that, although adults with skills of 200 were assigned to Level 1, because they could do 80% of the average tasks at that level, they could actually do 45% of the tasks at Level 2, 25% of those at Level 3, and even 15% (one in six) of those in Level 5. Adults with scores of 250 were assigned to Level 2, and it was implied that they could not perform more difficult tasks, even though they could do half (50%) of the tasks at Level 3, and one in five (20%) of the tasks at Level 5, the highest level of difficulty. But by being called Level 2 adults, all competence above that Level was (at least implicitly) denied to them.

Other Widely Used Standards Reduces Numbers of Adults in Levels 1 and 2 Dramatically. In a study of issues surrounding the setting of standards for adult literacy[37], Kolstad reported analyses showing that in the grade schools, the National Assessment of Educational Progress (NAEP) reports children's proficiency levels using a .65 probability of being able to perform the average task at the given level. Applying that standard to the

NALS prose scale data reduces the percentage of adults in Level 1 from 20 to 13 percent , and those in Level 2 drop from 27 to 19 percent. Altogether then, the percentage of adults below Level 3 drops from 47 to 32 percent, a 15 percent drop in adults considered marginally literate just by adopting for adults the same standard that is used for children in the K-12 school system!

The widely-used Comprehensive Adult Student Assessment System (CASAS) uses a probability of .50 to indicate a person's proficiency level. Applying the same standard to the NALS reduces the percentages in Level 1 from 20 to 9 percent, and in Level 2 from 27 to 13 percent. Combined, this reduces the percentage of adults below Level 3 by 25 percent, from 47 to 22 percent, or about one in five American adults in the lowest literacy level.

These new analyses about the rather arbitrary nature of standards for literacy led Kolstad to state, "A factor that has such a large impact on the results deserves a thorough understanding of the issues and debate over the standard to be adopted." This debate has yet to happen in adult education. Still, the question of "how good is good enough?" has been answered in practice by the National Governor's Association. It has established the national goal as getting all adults to score at Level 3 on the NALS scales. A daunting task given the fact that some 90 million adults are below the national standard.

The International Adult Literacy Survey (IALS)

A 1995 report from the Organization for Economic Co-operation (OECD) and Statistics Canada reported the results of testing of adults on NALS-type tests in different countries.[35] Table 5.3 shows the results for each of the three scales (prose, document, quantitative) in six of the countries that participated in the study. In detailed analyses across the various nations, major findings were similar in their trends to those found in the earlier NALS in the United States.

Table 5.3. Performance of adults aged 16-65 in six countries on prose, document and quantitative scales of the International Adult Literacy Survey (IALS).

Levels:	1			2			3			4/5		
Country	Prose	Doc.	Ont.	Prose	Doc.	Ont.	Prose	Doc.	Ont.	Prose	Doc.	Ont.
United Kingdm.	21.8	23.3	23.2	30.3	27.1	27.8	31.3	30.5	30.4	16.6	19.1	18.6
United States	20.7	23.7	21.0	25.9	25.9	25.3	32.4	31.4	31.3	21.1	19.0	22.5
Canada	16.6	18.2	16.9	25.6	24.7	26.1	35.1	32.1	34.8	22.7	25.1	22.2
Germany	14.4	09.0	06.7	34.2	32.7	26.6	38.0	39.5	43.2	13.4	18.9	23.5
Sweden	07.5	06.2	06.6	20.3	18.9	18.6	39.7	39.4	39.0	32.4	35.5	35.8
Poland	42.6	45.4	39.1	34.5	30.7	30.1	19.8	18.0	23.9	03.1	05.8	06.8

Test Score Conversions

Many adult programs do not use the NALS-type tests for measuring improvements in learning. In these programs, it may be desirable to identify correspondences among different adult literacy education tests so that cross-program comparisons can be made and progress toward the goal of having adult literacy students reach Level 3 of the NALS can be estimated from various tests. Following is a brief conversion chart for finding rough correspondences among the scale scores of the Comprehensive Adult Student Assessment System (CASAS), the reading grade level scores of the Adult Basic Learning Exam (ABLE) and the Tests of Adult Basic Education (TABE) and scale scores from the Tests of Adult Literacy Skills (TALS) scores, which are the same as the National Adult Literacy

Survey (NALS) scores. Scores in between those given can be obtained by drawing a graph with CASAS scores on the x axis and ABLE scores on the y axis and plotting the x and y data points in the chart. Then connect the plotted data points with a straight line. The same can be done for the other tests. Any correspondences needed can then be read off of the graph.

These data are based on studies in which the correlations between the CASAS and the other tests are in the .70 area. This leaves lots of room for variations in estimates. It is best to think of these scores as rough indicators of people's skills, perhaps as low, medium and higher levels. Keep in mind that we are not talking atomic clock accuracy when we measure adult literacy with any of these tests!

Conversion Chart

If CASAS score is 200, then ABLE score is 3.9; TABE score is 4.2, TALS score is 178.

If CASAS score is 215, then ABLE score is 6.6, TABE score is 7.0, TALS score is 229.

If CASAS score is 225, then ABLE score is 8.5, TABE score is 8.8, TALS score is 260.

If CASAS score is 230, then ABLE score is 9.4, TABE score is 9.8, TALS score is 279.

Conversion for CASAS to ABLE and TABE scores is described in Sticht [39] and for CASAS to TALS (or NALS) in Haney et. al.[40]. The estimate is from Table 6.1 of that report and is the average of four different methods given for converting CASAS to TALS scores.

Footnotes

[1] Public Law 100-297, Title III, Part A, Subpart 5, section 352: Evaluation.

[2] Federal Register, August 18, 1989, p. 34435.

[3] T. Sticht (1982). Evaluation of the "reading potential" concept for marginally literate adults. Washington, DC: Office of the Assistant Secretary of Defense (Manpower, Reserve Affairs, and Logistics).

[4] Regression to the mean also ocurs whenever a high scoring group has been separated from the total group and retested later on. In this case, the average score of the high scoring group will tend to decrease as it regresses to the mean of the total group.

[5] T. Sticht (1975). A program of Army functional job reading training: Development, implementation, and delivery systems (Final Report). HumRRO-FR-WD-(CA)-75-7. Alexandria, VA: Human Resources Research Organization.

[6] R. Glaser & D. Klaus (1962). Proficiency measurement: Assessing human performance. In: R. Gagne (Ed.), Psychological principles in system development. New York: Holt, Rinehart, and Winston.

[7] R. Taggart (1985, March). The Comprehensive Competencies Program: A summary of 1984 results. Washington, DC: Remediation and Training Institute.

[8] J. Davis, P. Williams, R. Stiles, J. Wise & P. Rickard (1984, April). CASAS: An effective measurement system for life skills. New Orlieans, LA: Paper presented at a meeting of the National Council on Measurement in Education.

[9] R. Bean, A. Byra, R. Johnson, & S. Lane (1988, July). Using curriculum-based measures to identify and monitor progress in an adult basic education program. Final Report. Pittsburgh, PA: University of Pittsburgh, Institute for Practice and Research in Education.

[10] L. Fuchs & S. Deno (1981). The relationship between curriculum-based mastery measures and standardized achievement tests in reading. (Research Report No. 57). Minneapolis, MN: University of Minnesota, Institute for Research on Learning Disabilities.

[11] J. Caylor & T. Sticht (1974, April). The problem of negative gain scores in the evaluation of reading programs. Chicago, IL: Paper presented at the annual meeting of the American Educational Research Association.

[12] T. Sticht (1988). Adult literacy education. In: E. Rothkopf (Ed.), Review of research in education. Volume 15. Washington, DC: American Educational Research Association.

[13] I. Kirsch & A. Jungeblut (1986). Literacy: Profiles of America's young adults. Princeton, NJ: Educational Testing Service.

[14] More can be learned about Item Response Theory (IRT) in a text and computer assisted instruction program: F. Baker (1985). The BASICS of item response theory. Portsmouth, NH: Heinemann Educational Books.

[15] The CASAS System (1989, November). GAIN Appraisal Program. III. Third Report. San Diego, CA: San Diego Community College District Foundation, Comprehensive Adult Student Assessment System.

[16] American Council on Education (1989). The Official GED practice Tests. Washington, DC: American Council on Education.

[17] T. Sticht, W. Armstrong, D. Hickey & J. Caylor (1987). Cast-off youth: Policy and training methods from the military experience. New York: Praeger.

[18] B. Gifford (Ed.) (1989). Test policy and the politics of opportunity allocation: The workplace and the law. Boston, MA: Kluver Academic Publishers.

[19] R. Pugsley (1987). National data update. Washington, DC: U.S. Department of Education, Paper presented at the Annual Conference, State Directors of Adult Education.

[20] D. Dixon, M. Vargo & D. Campbell (1987, July). Illiteracy in California: Needs, services & prospects. Palo Alto, CA: SRA Associates.

[21] K. Savage (undated, circa 1983). Teaching strategies for developing literacy skills in non-native speakers of English. Washington, DC: The National Institute of Education.

[22] P. Tirone (unpublished, circa 1988). Teaching and testing - can we keep our balance. New York: Literacy Assistance Center, Inc.

[23] S. Lytle, A. Belzer, K. Schultz & M. Vannozzi (1989). Learner-centered literacy assessment: An evolving process. In: A. Fingeret & P. Jurmo (Eds.), Participatory literacy education. San Francisco, CA: Jossey-Bass.

[24] R. Solorzano (1989, February). Analysis of learner progress from the first reporting cycle of the CALPEP field test. A report to the California State Librarian. Pasadena, CA: Educational Testing Service.

[25] See articles by M. Wolfe and S. Hill in the September 1989 special issue of Information Update published by the Literacy Assistance Center, Inc. of New York city. For an earlier application of portfolio-type assessment applied to adult education see R. Nickse (1980). Assessing life-skills competence. Belmont, CA: Pitman Learning, Inc.

[26] S. Hill (1989, September). Alternative assessment strategies: Some suggestions for teachers. Information Update, 6, pp. 7,9.

[27] M. Dick (1989, September). From the editor. Information Update, 6, pp. 1-2.

[28] R. Morris, L. Strumpf & S. Curnan (1988, May). Using basic skills testing to improve the effectiveness of remediation in employment and training programs for youth. Washington, DC: National Commission for Employment Policy.

[30] T. Sticht (1985). Understanding readers and their uses of texts. In: T. Duffy & R. Waller (eds.), Designing usable texts. New York: Academic Press.

[31] W. Grimes & W. Armstrong (unpublished, 1988-89). Test score data for the ABLE and CASAS survey of achievement tests. San Diego, CA: San Diego Community College District, Division of Continuing Education.

[32] Office of the Assistant Secretary of Defense (Manpower, Reserve Affairs and Logistics) (1982, March). Profile of American youth. Washington, DC.

[33] B. Waters, J. Barnes, P. Foley, S. Steinhaus & D. Brown (1988, October). Estimating the reading skills of military applicants: Development of an ASVAB to RGL conversion table. Final Report 88-82 (HumRRO FR-PRD-88-82). Alexandria, VA: Human Resources Research Organization.

[34]Kirsch, I., Jungeblut, A., Jenkins, L. & Kolstad, A. (1993, September). *Adult Literacy in America: A First Look at the Results of the National Adult Literacy Survey*. Washington, DC: U. S. Government Printing Office.

[35]Organization for Economic Co-operation and Development & Statistics Canada. (1995). *Literacy, Economy, and Society: Results of the First International Adult Literacy Survey*. Paris: Organization for Economic Co-operation and Development .

[36]National Education Goals Panel (1992). *The National Education Goals Report: Building a Nation of Learners*. Washington, DC: U. S. Government Printing Office, pp. 40-43.

[37]Kolstad, A. (1996, April). *The response probability convention embedded in reporting prose literacy levels from the 1992 National Adult Literacy Survey*. New York: Paper presented at the annual meeting of the American Educational Research Association.

[38]Bernstein, D., Roy, E., Srull, T & Wickens, C. (1988). *Psychology*. Boston: Houghton Mifflin.

[39]Sticht , T. (1990, January). *Testing and Assessment in Adult Basic Education and English as a Second Language*. Washington, DC: U. S. Department of Education, Adult Literacy Clearinghouse.

[40]Haney, W. et. al. (1994, October). *Calibrating scores on two tests of adult literacy: An equating study of the Test of Adult Literacy Skills (TALS) Document Test and the Comprehensive Adult Student Assessment System (CASAS) GAIN Appraisal Reading Test (Form 2)*. Chestnut Hill, MA: Boston College.

Appendix

Review of Tests for ABE and ESL Programs[a]

There are hundreds of standardized tests. Yet only a very few have been developed for use by ABE or ESL program providers.

This appendix provides reviews of eight standardized tests that are widely used by ABE and ESL programs. These tests were selected for review to include the most widely used group-administered, norm-referenced tests of adult basic skills (ABLE, TABE); the group-administered, competency-based tests of the CASAS; tests for ESL assessment (ESLOA; BEST; CASAS/ESL); tests that are used by volunteer adult literacy groups for individual testing in tutor-tutee arrangements (ESLOA; READ); and the GED Official Practice Test for indicating readiness for taking the GED high school equivalency examinations.

The information reported here for each test includes: the full name, commonly used acronym, and dates of publication; purpose; source; costs; description of skills assessed, reliability, validity, and types of scores that can be reported; and general comments. Notable strengths and weaknesses are high-lighted.

Reliability and validity coefficients are referred to as "low" when they are between 0 and .49, as "moderate" when between .50 and .79, and as "high" when equal to or greater than .80. When tests have different "levels" that means there are different tests for learners of different skill levels. The proper use of the appropriate level of test provides a more reliable estimate of learners' skills.

Final decisions about the use of any test should be made only after examining it carefully, reading its manual(s), and trying it with some students similar to those with whom it will be used.

Unless otherwise mentioned, the tests are suited to group administration, and the student test booklets are re-usable. The costs reported are for small orders and are only approximate, prices change over time; institutional or bulk order discounts are available from some publishers. Allow plenty of time when ordering materials. Order fulfillment normally takes 2-5 weeks unless special shipment and payment is specified. Errors in fulfilling orders are not uncommon.

[a]This appendix was written by Dr. Gregg Jackson in 1995, therefore the information is current as of that time. The reviews of tests are abstracts from more extensive reviews of 64 standardized tests and assessment instruments in a report prepared by Jackson for the Association for Community Based Education (ACBE) 1806 Vernon Street N.W., Washington, DC 20009, (202) 462-6333.

Adult Basic Learning Examination (ABLE, 1967-86)

Purpose: To measure several basic education skills of adults.

Source: The Psychological Corporation, Order Service Center, P.O. Box 839954, San Antonio TX 78283-3954; (800) 228-0752.

Costs: Learner test booklets cost $1.44; answer sheets cost $.50.

Description: There are sections on vocabulary, reading comprehension, spelling, language, number operations, and quantitative problem solving. There are three levels of the test, corresponding to skills commonly taught in grades 1-4, 5-8, and 9-12. There are two equivalent forms at each level for pre-and post-testing. A brief locator test is available to match the learners' skill levels to the appropriate level of test.

Reliability, Validity, and Scores: Test-retest reliability is not reported. Internal reliability has been high. Validity analyses show moderate correlations with the Stanford Achievement Test. Scores can be reported as scale scores, percentiles, stanines, and grade equivalents. Item response data are also reported. The norm data are based on 4,000 adults in 41 states and are reported separately for ABE/GED students, prisoners, vocational/technical students (only at Level 3), and a combination of all.

Comments: This is a 1986 revision of a test that has been widely used to evaluate the outcomes of adult basic education. The revision appears to be very responsive to several criticisms of prior tests used in adult basic education programs. The content and tone are adult. The reading passages are mostly about common everyday matters, and the questions tap not only literal comprehension, but also higher forms of comprehension. The mathematics word problems are representative of those many people encounter in daily life.

Ten of the items in the reading comprehension section of Level 1 (Form E) cannot be answered correctly without background knowledge that a moderate portion of adult learners will not possess or they require predicting what an imaginary person did in a given situation, and there is no way to know for sure. The "correct answer" presumes the imaginary person will act in the rational, safe, or common manner, but people do not always do so.

The Level 3 math section includes only a few very simple algebra and geometry problems. Some learners who score high may find themselves required to take remedial math when enrolling in technical schools and colleges.

This reviewer has extensive substantial experience in administering the reading comprehension and problem solving sections to adult literacy students. The students do not appear offended or antagonized by the test, they apply themselves and try to do well, and often perform somewhat better than their instructors had expected.

Basic English Skills Test (BEST, 1981-87)

Purpose: To assess speaking, listening, reading, and writing skills of low proficiency non-native English speakers.

Source: Center for Applied Linguistics, 1118 22nd Street N.W., Washington DC 20037; (202) 429-9292.

Costs: For the oral interview section, the administrator's picture cue books to which the learners respond cost $11.00 and answer sheets cost $.25; for the literacy skills section, the not re-usable learner test booklets and scoring sheets (together) cost $2.25.

Description: There are two sections. The oral interview section has 50 items and yields five scores for listening comprehension, pronunciation, communication, fluency, and reading/writing. It asks several personal questions, and then asks questions and gives the learners directions to follow in response to photographs, signs, a map, and some money placed on the table. The questions ask what are the people in the pictures doing, where is a specified object (the learner is to point to it), and what does a given sign mean. A few reading and writing items are included. The literacy skills section assesses reading and writing more thoroughly. There is only one level of the test. A second equivalent form of the test was recently made available.

Reliability, Validity, and Scores: Test-retest reliability is not reported in the manual. Internal reliability has been moderately high for the listening, communication, and fluency scores, and high for the total of the oral interview section. There are limited validity data. Learners assigned to seven ESL instructional levels, by means other than the BEST, were administered the BEST; the mean score of learners was substantially higher at each successive level. Though the test was administered to 987 ESL learners during its refinement, no norm data are reported in the manual. The manual describes "Student Performance Levels" for various total scores, but the basis for the specified levels is not given.

Comments: This test is adult in content and tone. The first section must be administered individually and to do so is moderately complex. Proper administration will require prior training and practice. The administration is paced and takes about 10 to 20 minutes. Most of the scoring of the first section is done as it is administered, not later from a tape recording. This saves time, but it can be distracting to the learner and sometimes even to the administrator. The scoring is judgmental and moderately complex, but after careful training inter-rater reliability has been high. A review of the test in Reviews of English Language Proficiency Tests (see Appendix B) described it as exciting, innovative, and valid, but time-consuming to use and lacking justification for the scoring system.

CASAS Adult Life Skills - Reading (1984-89)

Purpose: To assess a learner's ability to apply basic reading skills to common everyday life situations.

Source: CASAS, 8910 Claremont Mesa Blvd., San Diego, CA 92123; (619) 292-2900.

Costs: Special training by CASAS is required before using this test; write or call for fees and material costs.

Description: There is just one section of the test. Several levels are available, AA, A, B, C, suitable, respectively, for developmentally disabled and normal beginning, intermediate, and moderately advanced adult education learners. Level C is substantially easier than the GED test. There are two equivalent forms for each level. All CASAS tests are prepared from the CASAS item bank that now has 4,000 items. The bank permits quick and relatively inexpensive construction of customized tests for given objectives and difficulty levels. There are ready-made mathematics and English listening tests available.

Reliability, Validity, and Scores: Test-retest reliability is not reported. Internal reliability has been high. The manual and other publications sent to this reviewer do not indicate studies to validate the test against other measures of life-skills reading (though a moderate correlation of .70 was found in unpublished data for the ABLE and the CASAS reading test, see Appendix A, Table A-1 of this report). Raw scores are converted to CASAS scale scores; percentiles or grade equivalents are not reported. Data are presented for average entry, exit, and gains in programs throughout California over several years. Tables in the manual also indicate the specific objective measured by each item in the instruments.

Comments: This test is also referred to as the CASAS Survey Achievement Test. It is used widely in California by state-funded ABE and ESL programs, and it is also used elsewhere. The instrument is adult in content and tone. Virtually all of the reading materials are things that most adults would find very useful in everyday living. The content, however, is exclusively life-skill oriented. There are not items that use the kinds of reading material commonly found in popular magazines, newspapers, and books. Most of the items only assess literal reading comprehension. Few require inferences or evaluation.

Though CASAS is described as a competency-based assessment system, this reading test is not suited to assessing specific competencies. That is because the specified competencies are broad in scope and seldom measured by more than two items. For instance, in Form 31 of Level A, the competency of "interpret food packaging labels" is assessed by just one item, and the competency of "identify the months of the year and the days of the week" is assessed by only two items.

CASAS Adult Life Skills - Listening (1984-87)

Purpose: To assess English listening comprehension in common everyday life situations.

Source: CASAS, 8910 Claremont Mesa Blvd., San Diego, CA 92123; (619) 292-2900.

Costs: Special training by CASAS is required before using this test; write or call for fees and material costs.

Description: There are three levels, corresponding approximately to beginning, intermediate, and advanced ESL. There are two equivalent forms at each level. A cassette tape recording gives directions or asks a question, and the learner responds by selecting one of three alternative illustrations or sentences in a booklet. At the lowest level an example is: "Look at the pictures and listen [There are pictures of : a) a sheet of paper, b) a pencil, and c) a book]. What is the correct answer - A, B, or C? Give me a pencil. Is the answer A, B, or C?" At the low level, most items require no reading by the learners except of the letters "A," "B," and "C" used to designate the three pictures. At the intermediate level about half the items require reading at about the third grade level. At the high level, most of the items require reading at about the fifth grade level.

Reliability, Validity, and Scores: Reliability data are not reported in the materials examined. However, the test has been constructed in the same manner as several other CASAS tests that have had high internal reliability. Validity data are not provided, and may be questionable. As mentioned above, many of the items in the intermediate and high levels of the test require reading skills. It is likely that some learners who comprehend the spoken English directions and questions are unable to select the appropriate responses because of inadequate reading skills. This would be particularly true in ESL programs serving learners who are illiterate in their native language and those that focus exclusively on oral language instruction methods.

Comments: A commendable array of life-skills materials are included, and most people living in the United States would find it useful to master the listening comprehension that is measured by this test. The test is used widely in California, and is also used elsewhere.

This is one of the few tests of oral English skills that does not have to be administered to one learner at a time. But because it was designed for group administration, it only assesses passive or receptive, not interactive or conversational comprehension of oral English. It also does not assess the speaking of English. Some learners have comprehension skills substantially above their speaking skills.

English as a Second Language Oral Assessment (ESLOA, 1978-80)

Purpose: To efficiently measure the ability of non-native English speakers to understand and speak English.

Source: Literacy Volunteers of America, 5795 Widewaters Parkway, Syracuse NY 13214; (315) 445-8000.

Costs: The cue books cost $7.25; answer sheets cost $.04.

Description: The test is divided into four progressively more difficult levels. There is only one form of the test. The learner is judged as being at level 1, 2, 3, or 4, depending on how many levels he or she completes. At the first level, the student is shown drawings with three objects and asked questions like: "Where is the Box?" or "Which girl is walking?" The learner may respond orally or by pointing. At the second level, the learner is asked to answer simple questions and name illustrated objects. At the third level, the learner is shown drawings and asked questions such as: "What is he doing?" and "Where is she going?" The learner must respond orally, and is encouraged to use complete sentences. The learner is also orally given several sentences and asked to modify them in a specified manner, such as from statements to questions. At the fourth level, the learner is orally given sentences and asked to change them to different tenses, shown pictures and asked what is happening in them, and told of specific circumstances and asked what he or she would do in them. There also is an optional section that provides a simple means for judging spoken English in response to personal questions such as: "What television shows do you like? Why?"

Reliability, Validity, and Scores: The publisher does not have reliability or validity data. The cue book, which also serves as the manual, does not report any norm data. Lesson content is suggested for learners who score at each of the four specified levels.

Comments: This test is part of the materials prepared and distributed by Literacy Volunteers of America. Most items deal with commonly encountered objects and events, but few directly involve the activities that most occupy adults' lives - working, meal preparation, housekeeping, and child raising. The test focuses on beginning and intermediate English. People scoring at the highest level, Level 4, could easily have difficulty understanding and participating in conversational English.

The test must be administered individually. Administration is simple and is terminated when a learner misses more than a specified number of items on any of the four sections. There is no time limit; 10 to 20 minutes will usually be needed. Scoring is simple and quick.

GED Official Practice Tests (1987-88)

Purpose: To help learner's determine their readiness to take the GED tests.

Source: Prentice-Hall, 200 Old Tappan Road, Old Tappan NJ 07675; (800) 223-1360

Costs: Learner booklets cost $2.13; answer sheets cost $.25.

Description: There are five sub-tests. They cover writing, social studies, science, interpreting literature and the arts, and mathematics. The GED tests cover the same subjects, but are about twice as long as the practice tests. There is only one level of the practice tests, but there are two English forms for use in the U.S., one for use in Canada, and one form entirely in Spanish.

Reliability, Validity, and Scores: Test-retest reliability, using the two equivalent U.S. forms, has been high for each sub-test, when assessed with a large sample of high school seniors. Internal reliability, based on data from a sample of GED candidates was also high. The sub-test scores on the U.S. forms correlated moderately highly with the comparable GED test scores in a large sample of high school students. Validity coefficients for GED candidates are not reported. Raw scores are converted to the same standard scale scores as used for the GED tests. The manual also reports the subject area and cognitive skill covered by each multiple-choice item. This can be used to help diagnose particular weaknesses that a learner may have.

Comments: This test was developed by the same organization that prepares the GED tests, and in accordance with the same specifications used for those tests. The test is adult in content and tone. The orientation is generally middle class and academic, but that is appropriate since the same is true of the GED tests.

This is a good predictor of GED test performance, and probably the best available. But all tests have some measurement error. For a learner to be reasonably assured of passing the GED in a state that requires passing every sub-test, all his or her predictor sub-test scores should be at least 13 GED scale points above the minimum pass level. That requires getting about two-thirds of the items correct in each sub-test.

Though there is no sub-test that specifically assesses reading skills, this test requires much reading, with most of it at about the 11th grade level. The test also requires considerable application of critical thinking.

Scoring of the essay part of the writing sub-test is complex, requires prior training, and is time consuming. An explanation of the procedures and accompanying examples take 53 pages in the manual.

Reading Evaluation Adult Diagnosis (Revised) (READ, 1972-82)

Purpose: To assess learner's reading needs and progress.

Source: Literacy Volunteers of America, 5795 Widewaters Parkway, Syracuse NY 13214; (315) 445-8000.

Costs: The cue books cost $7.25. Answer sheets, suitable for two administrations to the same learner, cost $1.25.

Description: The test has three parts. The first part assesses sight word recognition - identifying words without the application of phonic analysis. The learner is shown lists of words and asked to read them aloud. The easiest list includes words like "he" and "big;" the most difficult list includes words like "family" and "arrive." The second part assesses word analysis - the application of phonics to unfamiliar words. Learners are asked to name the letters of the alphabet, pronounce consonants, and pronounce words that may be unfamiliar. The third part assesses reading or listening comprehension. The learner is asked to read aloud, and to listen to short passages and answer questions about them - who, what, where, and how? There are two approximately equivalent forms of Part 1 and Part 3 of the test; there is only one form of Part 2.

Reliability, Validity, and Scores: No data on reliability are reported in the cue book, which also serves as a manual, nor in the supplemental information requested from the publisher. No data on validity are reported in the cue book. Supplemental information sent by the publisher indicates that a prior version of this test, prepared by a different author, correlated moderately with the reading scores from the Adult Basic Learning Examination (ABLE). That does not indicate the validity of the current version. No norm data are reported. Implications for instruction are provided with each section of the test.

Comments: This test is part of the materials prepared and distributed by Literacy Volunteers of America. It is intended to be used for diagnosis and monitoring. The reading difficulty ranges up to only about grade 5. The short reading passages are generally adult in orientation, but they seem bland to this reviewer and may not be of high interest to many low-income adults.

The test must be administered individually. The instructions are moderately complex, sometimes awkward to comply with, and occasionally incomplete. The complexity is caused by the variety of different types of items, each with its own instructions; dividing instructions for a given exercise among non-contiguous pages; interspersing pre-test and post-test items in the display materials; and specifying various skip patterns depending on the learner's performance. There is no time limit and no indication of how long the test normally takes to administer. Manual scoring is moderately complex, but takes only a few minutes for each student.

Tests of Adult Basic Education - Forms 5 and 6 (TABE, 1957-87)

Purpose: To measure reading, writing, and mathematics achievement.

Source: Publisher's Test Service, CTB/McGraw-Hill, 2500 Garden Road, Monterey CA 93940; (800) 538-9547.

Costs: The learner test booklets cost $1.62; answer sheets cost $.43.

Description: There are seven sections measuring vocabulary, reading comprehension, language mechanics, language expression, spelling, mathematical calculation, and mathematical concepts/application. There are four levels corresponding in difficulty to grades 2-4, 4-6, 6-8, and 8-12. A locator test is available for matching learner skill levels to test levels. There are two equivalent forms at each level.

Reliability, Validity, and Scores: Test-retest reliability is not reported in the manuals. Internal reliability has been high. Limited validity data are reported in the manuals. The scores on the TABE have correlated moderately with comparable scores on the GED. Scores can be reported as scale scores, percentiles, stanines, and grade equivalents. The norm data are based on 6,300 learners in 223 institutions across the country. Norms are reported separately for adult basic education learners, adult offenders, juvenile offenders, and vocational/technical school enrollees. Data in the Norms Book also permit prediction of GED scores, but should be treated as rough estimates because of the moderate correlations between the TABE scores and the GED scores. The Test Coordinator's Handbook reports the knowledge and type of cognitive skill covered by each test item.

Comments: The TABE is one of the most widely used tests in adult basic education programs. It was thoroughly revised in 1986. All the items are new, the range of skill levels that can be assessed has been extended, and the specific skills that are measured are more finely divided and identified.

However, the lowest level of the test will be daunting and frustrating for most students with less than grade 3.0 skills. For instance, the first reading exercise uses a 150-word passage. Though the items are adult in content, they seem to this reviewer distinctly middle class and academic in orientation. Only a modest portion of them are about everyday events in the lives of low-income adults. For instance, in the grade 4-6 level booklet (Form 5M), only two of the eight reading passages are about experiences common to such learners. Of the 40 items on math concepts and application there is only one item on calculating the correct change for a given transaction, no item on the savings from bulk purchases, and no item on the total cost of a purchase with installment plan financing charges. The language sections are notable for focusing on paragraph construction as well as sentence structure.

This test assesses an unusually broad range of skills. Therefore, giving the full TABE takes about 4.5 hours. For this reason, many programs use only one or two sections for pre- and post-testing.